

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 4e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGLEFSQAKIS 18  
:|||||:|||||:|  
DB 822 ILPPPRKMGLEFSQSKIS 839

RESULT 10  
POLG\_HPAV1 STANDARD; PRT; 341 AA.  
AC P13672;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last annotation update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP3; Core protein P2A] (Fragment).  
OS Hepatitis A virus (strain LCD-1).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12093;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89263805; PubMed=2542903;  
RA Andonov A.P., Lau P., Chaudhary R.;  
RT "Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus (HAV).";  
RL Nucleic Acids Res. 17:3594-3594(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X14666; CA32794.1; -.  
DR PIR: S04137; S04137.  
KW Polypeptide; Coat protein; Core protein.  
FT NON\_TER 1 1  
FT CHAIN <1 1 COAT PROTEIN VP3 (1C).  
FT CHAIN 2 340 COAT PROTEIN VP1 (1D).  
FT CHAIN 341 >341 CORE PROTEIN P2A.  
FT NON\_TER 341 341  
SQ SEQUENCE 341 AA; 38003 MW; 066918289BFL126D5 CRC64;

Query Match 74.3%; Score 75; DB 1; Length 341;  
Best Local Similarity 93.3%; Pred. No. 4.8e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGLEFSQAKIS 15  
:|||||:|||||:|  
DB 327 VLPPPRKMGLEFSQSKIS 341

RESULT 11  
DPYD\_HUMAN STANDARD; PRT; 1025 AA.  
ID Q12882; Q12882; Q12882; Q12882;  
AC Q12882; Q12882; Q12882; Q12882;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Dihydropyrimidine dehydrogenase [NADP+] precursor (PC 1.3.1.2) (DPD)  
DE (DHPHase) (Dihydropyrimidine dehydrogenase) (Dihydrothymine  
DE dehydrogenase).

GN DPYD.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94365020; PubMed=8083224;  
RA Yokota H., Fernandez-Salguero P., Puruya H., Lin K., McBride O.W.,  
RA Podschun B., Schmeckler K.D., Gonzalez F.J.;  
RT "cDNA cloning and chromosome mapping of human dihydropyrimidine  
RT dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and  
RT congenital thymine uraciluria.";  
RL J. Biol. Chem. 269:23192-23196(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97280676; PubMed=9135003;  
RA Johnson M.R., Wang K., Tillmans S., Albin N., Diasio R.B.;  
RT "Structural organization of the human dihydropyrimidine dehydrogenase  
RT gene.";  
RL Cancer Res. 57:1660-1663(1997).  
RN [3]  
RP SEQUENCE OF 581-635 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97047101; PubMed=8892022;  
RA Vreken P., van Kullenburg A.B.P., Meinsma R., Smit G.P.A.,  
RA Bakker H.D., de Abreu R.A., van Gennip A.H.;  
RT "A point mutation in an invariant splice donor site leads to exon  
RT skipping in two unrelated Dutch patients with dihydropyrimidine  
RT dehydrogenase deficiency.";  
RL J. Inher. Metab. Dis. 19:645-654(1996).  
RN [4]  
RP CHARACTERIZATION, AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=92381021; PubMed=1512248;  
RA Lu Z.-H., Zhang R., Diasio R.B.;  
RT "Purification and characterization of dihydropyrimidine dehydrogenase  
RT from human liver.";  
RL J. Biol. Chem. 267:17102-17109(1992).  
RN [5]  
RP VARIANTS ARG-29; TRP-235 AND HIS-886.  
RX MEDLINE=98102836; PubMed=9439663;  
RA Vreken P., van Kullenburg A.B.P., Meinsma R., van Gennip A.H.;  
RT "Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and  
RT expression of missense mutations C29R, R886H and R235W.";  
RL Hum. Genet. 101:333-338(1997).  
RN [6]  
RP VARIANTS ARG-29; TRP-235 AND HIS-886.  
RX MEDLINE=97411371; PubMed=9266349;  
RA Vreken P., van Kullenburg A.B.P., Meinsma R., van Gennip A.H.;  
RT "Identification of novel point mutations in the dihydropyrimidine  
RT dehydrogenase gene.";  
RL J. Inher. Metab. Dis. 20:335-338(1997).  
RN [7]  
RP FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE  
RP REDUCTION OF URACIL AND THYMINE. ALSO INVOLVED THE DEGRADATION OF  
RP THE CHEMOTHERAPEUTIC DRUG 5-FLUOROURACIL.  
CC CC  
CC -1- CATALYTIC ACTIVITY: 5.6-dihydrouracil + NADP(+) -> uracil + NADPH.  
CC -1- COFACTOR: TWO EACH OF FAD AND FMN. ALSO CONTAINS TWO 4FE-4S  
CC CLUSTERS. CONTAINS APPROXIMATELY 33 IRON ATOMS PER MOLECULE.  
CC -1- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF  
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE  
CC FORMATION OF BETA-ALANINE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES WITH GREATEST ACTIVITY  
CC FOUND IN LIVER AND PERIPHERAL BLOOD MONONUCLEAR CELLS.  
CC -1- DISEASE: DEFECTS IN DPYD ARE THE CAUSE OF HEREDITARY THYMINE-  
CC URACILURIA (ALSO KNOWN AS FAMILIAL PYRIMIDINEMIA), A DISEASE  
CC CHARACTERIZED BY PERSISTENT URINARY EXCRETION OF EXCESSIVE AMOUNTS  
CC OF URACIL, THYMINE AND 5-HYDROXYMETHYLMTHYURACIL. PATIENTS SUFFERING  
CC FROM THIS DISEASE SHOW A SEVERE REACTION TO THE ANTICANCER DRUG 5-  
CC FLUOROURACIL. THIS REACTION INCLUDES STOMATITIS, LEUKOPENIA,

DR EMBL: X15461; CAA33490.1; -  
DR PIR: A30470; GNNYSA.  
DR PIR: S04885; S04885.  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 27 COAT PROTEIN VP4 (PIA).  
FT CHAIN 28 249 COAT PROTEIN VP2 (PIB).  
FT CHAIN 250 495 COAT PROTEIN VP3 (PIC).  
FT CHAIN 496 795 COAT PROTEIN VP1 (PID).  
FT CHAIN 796 984 CORE PROTEIN P2A.  
FT CHAIN 985 1091 CORE PROTEIN P2B.  
FT CHAIN 1092 1426 CORE PROTEIN P2C.  
FT CHAIN 1427 1498 PROBABLE PROTEIN 3A.  
FT CHAIN 1499 1521 PROBABLE PROTEIN 3B.  
FT CHAIN 1522 1741 PROBABLE PROTEIN 3C.  
FT CHAIN 1742 2230 RNA-DIRECTED POLYMERASE 3D.  
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match Best Local Similarity 95.0%; Score 96; DB 1; Length 2230;  
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPRRKKKGLFSQAKISLF 20  
Db 827 ILPPRRKKKGLFSQAKISLF 846  
:|||||:|||||

RESULT 8  
POLG\_HPAVM STANDARD; PRT; 2227 AA.  
AC P31901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MBB).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Waisel T., Klehn R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
RT virus (isolate MBB)."  
RL Virus Res. 8:153-161(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M20273; AAA45474.1; -

DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).  
FT CHAIN 837 980 CORE PROTEIN P2A.  
FT CHAIN 981 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.  
FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.  
FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.  
FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match Best Local Similarity 94.1%; Score 95; DB 1; Length 2227;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPRRKKKGLFSQAKISLF 20  
Db 823 VLPPRRKKKGLFSQAKISLF 842  
:|||||:|||||

RESULT 9  
POLG\_HPAVT STANDARD; PRT; 839 AA.  
AC P31788;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core protein  
DE P2A] (Fragment).  
OS Simian hepatitis A virus (strain CY-145).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=31707;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9131421; PubMed=1649902;  
RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
RT "Sequence analysis of a new hepatitis A virus naturally infecting  
RT cynomolgus macaques (Macaca fascicularis)."  
RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M59286; AAA45473.1; -  
DR PIR: J01180; GNNYS2.  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 7 COAT PROTEIN VP1 (PID).  
FT CHAIN ? 839 CORE PROTEIN P2A.  
FT NON\_TER 839 839  
SQ SEQUENCE 839 AA; 93825 MW; 2CACC4BD1E192DBC CRC64;

CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).  
-----  
DR EMBL: M59809: AAA45469.1: -.  
DR MEROPS: C03.005: -.  
DR InterPro: IPR000605: RNA\_helicase.  
DR InterPro: IPR001205: RNA\_pol\_P3D.  
DR Pfam: PF00680: RNA\_dep\_RNA\_pol: 1.  
DR Pfam: PF00910: RNA\_helicase: 1.  
KM Polypeptide: Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; ribol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 COAT PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;  
Best Local Similarity 95.0%; Pred. No. 1.2e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20  
|||||  
DB 823 VLPPRRKMKGLFSQAKISLF 842

RESULT 6  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination".  
RT J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
-----  
DR EMBL: M59808: AAA45467.1: -.  
DR MEROPS: C03.005: -.  
DR InterPro: IPR000605: RNA\_helicase.

DR InterPro: IPR001205: RNA\_pol\_P3D.  
DR Pfam: PF00680: RNA\_dep\_RNA\_pol: 1.  
DR Pfam: PF00910: RNA\_helicase: 1.  
KM Polypeptide: Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; ribol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 COAT PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396CD68 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;  
Best Local Similarity 95.0%; Pred. No. 1.2e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20  
|||||  
DB 823 VLPPRRKMKGLFSQAKISLF 842

RESULT 7  
POLG\_HPAVS STANDARD; PRT; 2230 AA.  
AC P14553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311420; PubMed=1649901;  
RA Taarev S.A., Emerson S.U., Balaayan M.S., Ticehurst J.R.,  
RA Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
RT structure and growth in cell culture with other HAV strains".  
RT J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE=89232168; PubMed=2541023;  
RA Balaayan M.S., Kusov Y.Y., Andjapardize A.G., Taarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
RT and simian hepatitis A viruses".  
RT FEBS Lett. 247:425-428(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
-----  
DR EMBL: D00924: BAA00766.1: -.

```
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: K02990; AAA45472.1; -.
CC PIR: A03903; GNNYHR.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol.1.
CC DR Pfam: PF00910; RNA_helicase.1.
CC KW Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC FT CHAIN 1 23
CC FT CHAIN 24 245
CC FT CHAIN 246 491
CC FT CHAIN 492 836
CC FT CHAIN 837 980
CC FT CHAIN 1077 1422
CC FT CHAIN 1423 1484
CC FT CHAIN 1485 1507
CC FT CHAIN 1508 1678
CC FT CHAIN 1679 2227
CC SO SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20
DB 823 VLPPRRKMKGLFSQAKISLF 842

RESULT 4
POLG_HPAV2 STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M59810; AAA45468.1; -.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol.1.
CC DR Pfam: PF00910; RNA_helicase.1.
CC KW Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC FT CHAIN 1 23
CC FT CHAIN 24 245
CC FT CHAIN 246 491
CC FT CHAIN 492 794
CC FT CHAIN 795 900
CC FT CHAIN 901 1087
CC FT CHAIN 1088 1422
CC FT CHAIN 1423 1495
CC FT CHAIN 1496 1518
CC FT CHAIN 1519 1737
CC FT CHAIN 1738 2226
CC SO SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;
Best Local Similarity 95.0%; Pred. No. 1.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20
DB 823 VLPPRRKMKGLFSQAKISLF 842

RESULT 5
POLG_HPAV4 STANDARD; PRT; 2226 AA.
AC P26581;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 43c).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12095;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```



FT CHAIN 837 >852 CORE PROTEIN P2A.  
FT NON\_TER 852  
SQ SEQUENCE 852 AA: 95563 MW: 73D3ED0AD532820E CRC64;

Query Match 100.0%; Score 101; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 6.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VLPPPRKMGLEFSQAKISLF 20  
Db 823 VLPPPRKMGLEFSQAKISLF 842

RESULT 2  
POLG\_HPAVL STANDARD: PRT: 2227 AA.  
AC P06617; P066443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain HM-175).  
OC Hepatitis A virus positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatitis A virus.  
OX NCBI\_Taxid=12098;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WILD TYPE;  
RX MEDLINE=87061253; PubMed=3023706;  
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A., Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses.";  
RT J. Virol. 61:50-59(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATTENUATED;  
RX MEDLINE=87175701; PubMed=3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M., Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with wild-type virus.";  
RT Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RN [3]  
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE=85166289; PubMed=2984684;  
RA Baroudy B.M., Ticehurst J.R., Mele T.A., Matzel J.V. Jr., Purcell R.H., Feinstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA polymerase.";  
RT Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
RN [4]  
RP SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1- PM: SPECIFIC ENZYATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT SHOWN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: M14114; AAA45475.1; -;  
DR EMBL: M14707; AAA45465.1; -;  
DR EMBL: M14707; AAA45466.1; ALT\_INIT.  
DR EMBL: M16632; AAA45471.1; -;  
DR PIR: A25981; GNNYMK.  
DR PIR: A25914; GNNYMK.  
DR PIR: A03905; A03905.  
DR MEROPS: C03.005; -;  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980.  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT VARIANT 77 77  
FT VARIANT 764 764  
FT VARIANT 821 821  
FT VARIANT 1052 1052  
FT VARIANT 1062 1062  
FT VARIANT 1118 1118  
FT VARIANT 1151 1151  
FT VARIANT 1163 1163  
FT VARIANT 1277 1277  
FT VARIANT 1500 1500  
FT VARIANT 1805 1805  
FT VARIANT 1930 1930  
SQ SEQUENCE 2227 AA: 251506 MW: 01E225E7AEB740A6 CRC64;  
K -> R (IN ATTENUATED STRAIN)  
E -> V (IN ATTENUATED STRAIN)  
N -> S (IN ATTENUATED STRAIN)  
A -> V (IN ATTENUATED STRAIN)  
G -> A (IN ATTENUATED STRAIN)  
K -> M (IN ATTENUATED STRAIN)  
E -> K (IN ATTENUATED STRAIN)  
F -> S (IN ATTENUATED STRAIN)  
V -> I (IN ATTENUATED STRAIN)  
H -> Y (IN ATTENUATED STRAIN)  
D -> N (IN ATTENUATED STRAIN)  
S -> T (IN ATTENUATED STRAIN).

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VLPPPRKMGLEFSQAKISLF 20  
Db 823 VLPPPRKMGLEFSQAKISLF 842  
RESULT 3  
POLG\_HPAVL STANDARD: PRT: 2227 AA.  
AC P06641;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain LA).  
OC Hepatitis A virus positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatitis A virus.  
OX NCBI\_Taxid=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85190549; PubMed=2986127;  
RA Marjarian R., Caput D., Gee W.W., Potter S.J., Renard A., Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
RT Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
RN [2]  
RP SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:10:07 ; Search time 59.43 Seconds

(without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432a-42

Perfect score: 101

Sequence: 1 VLPPPRKMGLEFSQAKISLF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : SwissProt\_40:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	101	100.0	852	1	P06442 hepatitis a
2	101	100.0	2227	1	P08617 hepatitis a
3	101	100.0	2227	1	P06441 hepatitis a
4	96	95.0	2226	1	P06441 hepatitis a
5	96	95.0	2226	1	P26580 hepatitis a
6	96	95.0	2226	1	P26581 hepatitis a
7	96	95.0	2226	1	P26582 hepatitis a
8	95	94.1	2227	1	P14553 simian hepa
9	84	83.2	839	1	P13901 hepatitis a
10	75	74.3	341	1	P06441 hepatitis a
11	52.5	52.0	1025	1	P06441 hepatitis a
12	45.5	45.0	1025	1	P06441 hepatitis a
13	44	43.6	1025	1	P06441 hepatitis a
14	44	43.6	1025	1	P06441 hepatitis a
15	43	42.6	389	1	P06441 hepatitis a
16	43	42.6	389	1	P06441 hepatitis a
17	41	40.6	389	1	P06441 hepatitis a
18	41	40.6	389	1	P06441 hepatitis a
19	41	40.6	389	1	P06441 hepatitis a
20	41	40.6	389	1	P06441 hepatitis a
21	40	39.6	372	1	P06441 hepatitis a
22	40	39.6	372	1	P06441 hepatitis a
23	40	39.6	372	1	P06441 hepatitis a
24	40	39.6	372	1	P06441 hepatitis a
25	40	39.6	372	1	P06441 hepatitis a
26	40	39.6	372	1	P06441 hepatitis a
27	40	39.6	372	1	P06441 hepatitis a
28	40	39.6	372	1	P06441 hepatitis a
29	40	39.6	372	1	P06441 hepatitis a
30	40	39.6	372	1	P06441 hepatitis a
31	40	39.6	372	1	P06441 hepatitis a
32	40	39.6	372	1	P06441 hepatitis a
33	39.5	39.1	1040	1	P06441 hepatitis a

34	39	38.6	204	1	RS4_TREPA	083328 treponema p
35	39	38.6	347	1	E13A_SOYBN	003773 glycine max
36	39	38.6	427	1	TRB1_AERPE	094875 aeropyrum p
37	39	38.6	462	1	MDR8_MOUSE	091m98 mus musculus
38	39	38.6	615	1	CKN_RAT	094011 rattus norv
39	39	38.6	631	1	CKN_MOUSE	060806 mus musculus
40	39	38.6	646	1	CKN_HUMAN	094484 homo sapien
41	39	38.6	808	1	POLG_HPARG	002381 hepatitis a
42	39	38.6	1103	1	DPOD_RAT	054747 rattus norv
43	39	38.6	1248	1	CYAB_RAT	P40146 rattus norv
44	39	38.6	1249	1	CYAB_MOUSE	P97470 mus musculus
45	39	38.6	1251	1	CYAB_HUMAN	P40145 homo sapien
46	39	38.6	1387	1	RSCC_RAT	008774 rattus norv
47	39	38.6	1585	1	ARP8_YEAST	P35522 saccharomyc
48	38.5	38.1	881	1	YD03_ARCFU	012386 saccharomyc
49	38	37.6	49	1	MEK2_CHICK	028966 archaeoglob
50	38	37.6	398	1	FXP3_MOUSE	090891 gallus galli
51	38	37.6	429	1	FXP3_MOUSE	099186 mus musculus
52	38	37.6	465	1	HN4A_MOUSE	P41235 homo sapien
53	38	37.6	465	1	HN4A_MOUSE	P49698 mus musculus
54	38	37.6	465	1	HN4A_MOUSE	P22449 rattus norv
55	38	37.6	491	1	YA29_SCHPO	009700 sciliosasch
56	38	37.6	496	1	MEIA_TREPA	056342 treponema p
57	38	37.6	516	1	RORG_MOUSE	P51450 mus musculus
58	38	37.6	752	1	CLPE_STRPN	P35594 streptococc
59	38	37.6	1019	1	DLP2_HUMAN	0991a6 homo sapien
60	38	37.6	1089	1	DLP2_RAT	P97837 rattus norv

## ALIGNMENTS

RESULT	ID	STANDARD:	PRT:	852 AA.
1	POLG_HPARG			
AC	P06442: 083741: 083742:			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).			
OS	Hepatitis A virus (strain CR326).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.			
OX	NCBI_TaxID=12097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85185648; PubMed=2985793;			
RA	Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,			
RA	Young A., Mitra S.W.;			
RT	"Molecular cloning and partial sequencing of hepatitis A viral cDNA."			
RL	J. Virol. 54:247-255(1985).			
CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,			
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,			
CC	VP3, AND VP4.			
CC	-1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD NATURE PROTEINS.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-slb.ch/announce/">http://www.isb-slb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-slb.ch">license@isb-slb.ch</a> ).			
CC	-----			
DR	EMBL, M10033; AAA45470.1; -			
DR	PIR, A03904; GNNYHA.			
KW	POLYPEPTIDE; Coat protein; Core protein.			
FT	CHAIN 1 23			
FT	CHAIN 24 245			
FT	CHAIN 246 491			
FT	CHAIN 492 836			
FT	CHAIN COAT PROTEIN VP1 (P1D).			



A:Molecule type: DNA  
A:Residues: 1-105 <KAM>  
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80240.1; PID:01044026; PID:9510  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APEI250

Query Match 42.1%; Score 42.5; DB 2; Length 105;  
Best Local Similarity 61.1%; Pred. No. 7.6;  
Matches 11; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

OY 1 VLPPRRKMGLEFSQAKIS 18  
||| | | | : | | | |  
DB 36 VLPPRRAM-SMASOGLIS 52

RESULT 23  
S69307

Probable membrane protein YLR294C - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L8003.19-a

C:Species: Saccharomyces cerevisiae

C>Date: 20-Jul-1996 #sequence\_revision 23-Aug-1996 #text\_change 05-Nov-1999

C:Accession: S69307

R:Pauley, A.  
Submitted to the EMBL Data Library, November 1994

A:Description: The sequence of S. cerevisiae cosmid 8003.

A:Reference number: S50366

A:Accession: S69307

A:Molecule type: DNA

A:Residues: 1-109 <PAUD>

A:Cross-references: EMBL:U17243; NID:9596030; PID:92340967; GSPDB:GN00012; MIPS:YLR294C

C:Genetics:

A:Gene: MIPS:YLR294C

A:Map position: 12R

C:Superfamily: Saccharomyces probable membrane protein YLR294C

C:Keywords: transmembrane protein

F:77-93/Domain: transmembrane #status predicted <TMM>

Query Match 41.6%; Score 42; DB 2; Length 109;  
Best Local Similarity 50.0%; Pred. No. 9.6;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 5 PRKMGLEFSQAKISLF 20  
| | | | | : | | | |  
DB 63 PRKMGLEFSQAKISLF 78

RESULT 24  
S02053

hypothetical protein (D10 5' region) - phage T5

C:Species: phage T5

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 08-Oct-1999

C:Accession: S02053

R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.

Nucleic Acids Res. 16, 10353-10354, 1988

A:Title: The nucleotide sequence of the region of bacteriophage T5 early genes D10-D15.

A:Reference number: S01931; MUID:8057468

A:Accession: S02053

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-165 <KAL>

A:Cross-references: EMBL:X12930; NID:915407; PIDN:CAA31397.1; PID:9579170

C:Genetics:

A:Start codon: GTC

C:Superfamily: phage T5 hypothetical protein (D10 5' region)

Query Match 41.6%; Score 42; DB 2; Length 165;  
Best Local Similarity 53.3%; Pred. No. 15;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 PRKMGLEFSQAKISLF 20  
| | | | | : | | | |  
DB 38 RLPCTFSORRLILF 52

RESULT 25  
T02335

ribosomal protein L18a, cytosolic - Arabidopsis thaliana

N:Alternate names: protein F13p17.31

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001

C:Accession: T02335; B84757

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, July 1998

A:Description: Arabidopsis thaliana chromosome II BAC F13p17 genomic sequence.

A:Reference number: Z14657

A:Accession: T02335

A:Status: translated from GR/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-178 <ROU>

A:Cross-references: EMBL:AC004481; NID:93337347; PIDN:AMC27421.1; PID:93337376

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanRken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: B84757

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <STO>

A:Cross-references: GB:AE002093; NID:93128228; PIDN:AMC26708.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g34480; F13p17.31

A:Map position: 2

A:Introns: 5/3; 44/2; 65/3

C:Superfamily: rat ribosomal protein L18a

C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 41.6%; Score 42; DB 2; Length 178;  
Best Local Similarity 45.0%; Pred. No. 17;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 VLPPRRKMGLEFSQAKISLF 20  
| | | | | : | | | |  
DB 158 VRPPSRKLTYYKANKPNLF 177

Search completed: June 16, 2002, 00:05:19  
Job time: 793 sec

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: A12534  
R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Status: preliminary  
A:Accession: A12534  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-189 <KUR>  
A:Cross-references: GB:AP003602; PIDN:BA677207.1; PID:q17134649; GSPDB:GN00181  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1564  
A:Genome: plasmid

Query Match 42.6%; Score 43; DB 2; Length 189;  
Best Local Similarity 55.6%; Pred. No. 12;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 LPPRKMKGLEFSQAKISL 19  
| | | | | : | | |  
Db 162 LTPPLKMKMKLMQRMKILL 179

RESULT 19  
D75008  
methanol dehydrogenase regulatory protein (moxr-2) PAB1296 - Pyrococcus abyssi (strain C)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: D75008  
R:anonymous, Genoscope  
Submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: D75008  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-309 <KAW>  
A:Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50506.1; PID:9545902  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1296  
C:Superfamily: methanol dehydrogenase regulatory protein

Query Match 42.6%; Score 43; DB 2; Length 309;  
Best Local Similarity 81.8%; Pred. No. 21;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 RKMKGLEFSQAK 16  
| | | | | : | | |  
Db 185 RKDKGLFSEAK 195

RESULT 20  
T49919  
hypothetical protein F17114.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49919  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, I.  
Submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24490  
A:Accession: T49919  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-379 <BEV>  
A:Cross-references: EMBL:AL353994; GSPDB:GN00063; ATSP:F17114.20  
A:Experimental source: cultivar Columbia; BAC clone F17114

C:Genetics:  
A:Gene: ATSP:F17114.20  
A:Map position: 5  
A:Introns: 119/1; 146/1; 222/3; 256/2; 327/2

Query Match 42.6%; Score 43; DB 2; Length 379;  
Best Local Similarity 77.8%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 PPPRKMKG 11  
| | | | | : | | |  
Db 32 PPPRKMMS 40

RESULT 21  
S41748  
heat shock protein dnaJ - Methanosarcina mazei

C:Species: Methanosarcina mazei  
C:Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 20-Aug-1999  
C:Accession: S41748; S41150  
R:Conway de Macario, E.  
Submitted to the EMBL Data Library, April 1992  
A:Reference number: S41748  
A:Accession: S41748  
A:Molecule type: DNA  
A:Residues: 1-389 <CON>

A:Cross-references: EMBL:X60265; NID:948938; PIDN:CAA42813.1; PID:948940  
A:Experimental source: strain S-6  
R:Macario, A.J.L.; Dugan, C.B.; Clarens, M.; Conway de Macario, E.  
Nucleic Acids Res. 21, 2773, 1993  
A:Title: dnaJ in Archaea.  
A:Reference number: S41150; MUID:93324351  
A:Accession: S41150  
A:Status: nucleic acid sequence not shown

A:Molecule type: DNA  
A:Residues: 4-208 <MAC>  
A:Cross-references: EMBL:X60265; NID:948938; PIDN:CAA42813.1; PID:948940  
A:Experimental source: strain S-6  
C:Genetics:  
A:Gene: dnaJ  
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
C:Keywords: DNA replication; heat shock; molecular chaperone; stress-induced protein  
F:6-70/Domain: dnaJ amino-terminal homology <DMJ>  
F:77-104/Region: G/F motif  
F:144-151/Region: CXXCXGXG repeat  
F:161-168/Region: CXXCXGXG repeat  
F:187-194/Region: CXXCXGXG repeat  
F:201-208/Region: CXXCXGXG repeat

Query Match 42.6%; Score 43; DB 2; Length 389;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 PRKMKGLEFSQAK 16  
| | | | | : | | |  
Db 373 PRKSGLEFKVK 384

RESULT 22  
B72598  
hypothetical protein APE1250 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: B72598  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339  
A:Accession: B72598  
A:Status: preliminary



Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>  
A:Cross-references: EMBL:X14666; NID:962301; PIDN:CAA32794.1; PID:94377576  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; polypeptide  
F:2-340/Product: coat protein ID (VP1) #status predicted <MAN>

Query Match 74.3%; Score 75; DB 2; Length 341;  
Best Local Similarity 93.3%; Pred. No. 0.0001;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQA 15  
|||||:|||||:  
Db 327 VLPPRRKMKGLFSQS 341

RESULT 10  
A54718  
diHydroPyrimidine dehydrogenase (NADP+) (EC 1.3.1.2) - human  
C:Species: Homo sapiens (man)  
C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 17-Mar-2000  
C:Accession: A54718  
R:Yokota, H.; Fernandez-Salguero, P.; Furuya, H.; Lin, K.; McBride, O.W.; Podschun, B.;  
J. Biol. Chem. 269, 23192-23196, 1994  
A:Title: cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, a  
A:Reference number: A54718; MUID:94365020  
A:Accession: A54718  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1025 <YOK>  
A:Cross-references: GB:U09178; NID:9558304; PID:9558305  
A:Note: authors translated the codon GGC for residue 748 as Asp, and TTA for residue 101  
C:Genetics:  
A:Gene: GDB:DPYD  
A:Cross-references: GDB:364102; OMIM:274270  
A:Map position: 1p22-1p22  
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h  
C:Keywords: 4Fe-4S; flavoprotein; homodimer; iron-sulfur protein; NADP; oxidoreductase  
F:946-1005/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 2;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPRRKMKGLFSQAISLF 20  
|||||:|||||:  
Db 176 LPPRRKMKSEAYS-AKIALF 193

RESULT 11  
D96838  
unknown protein T21F11.5 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96838  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzalla,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719  
A:Accession: D96838  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-414 <STO>  
A:Cross-references: GB:AE005173; NID:96730725; PIDN:AAF27115.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T21F11.5  
A:Map position: 1

Query Match 46.5%; Score 47; DB 2; Length 414;  
Best Local Similarity 56.2%; Pred. No. 6.2;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPRRKMGLFSQAIS 18  
|||||:|||||:  
Db 106 PPRRKMGLFSQAIS 121

RESULT 12  
AG1062  
bacteriophage gene regulatory protein SRY4826 [imported] - Salmonella enterica subsp.  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AG1062  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
, S.; Moule, S.; O'Gea, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AG1062  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06948.1; PID:91505595; GSPDB:GN00176  
C:Genetics:  
A:Gene: SRY4826

Query Match 44.6%; Score 45; DB 2; Length 79;  
Best Local Similarity 63.6%; Pred. No. 2.1;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 PPRRKMGLFS 13  
|||||:|||||:  
Db 64 PPRRKMGLFS 74

RESULT 13  
E86423  
probable 60S ribosomal protein L18A - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
C:Accession: E86423  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzalla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E86423  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-321 <STO>

OY 1 VLPPPKMKGLFSQAKISLF 20  
DB 823 VLPPPKMKGLFSQAKISLF 842

## RESULT 6

GNNYSA

genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tsarev, S.A.  
submitted to JRPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:9222597; PIDN:BA00766.1; PID:9222598  
R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Tischenst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure an  
A:Reference number: J01080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhko  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:961971; PIDN:CAA33490.1; PID:9930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhko  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-493/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <C3A>  
F:1499-1521/Product: protein 3B #status predicted <C3B>  
F:1522-1741/Product: protein 3C #status predicted <C3C>  
F:1742-2230/Product: protein 3D #status predicted <C3D>

Query Match 95.0%; Score 96; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 2.5e-07;

Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPKMKGLFSQAKISLF 20  
DB 827 ILPPPKMKGLFSQAKISLF 846

## RESULT 7

GNNYHB

genome polyprotein - human hepatitis A virus (strain MBB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
Vpg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48); protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303

R:Paul, A.V.; Tada, H.; von der Helm, K.; Wiesel, T.; Klein, R.; Wimmer, E.; Deinhard  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso  
A:Reference number: J50303; MUID:88045071  
A:Accession: J50303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAD>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: coat protein 2A #status predicted <P2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: genome-linked protein vpg #status predicted <VP6>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 94.1%; Score 95; DB 1; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 3.6e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLPPPKMKGLFSQAKISLF 20  
DB 823 VLPPPKMKGLFSQAKISLF 842

## RESULT 8

GNNY52

genome polyprotein - simian hepatitis A virus (strain Cy-145) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat prot  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (cynomolgus macaque)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: J01180  
R:Nahman, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus  
A:Reference number: J01180; MUID:91311421  
A:Accession: J01180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NA1>  
A:Cross-references: GB:M59286; NID:9329599; PIDN:AAA45473.1; PID:9555083  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP0>  
F:246-491/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 8.5e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPKMKGLFSQAKIS 18  
DB 822 ILPPPKMKGLFSQAKIS 839

## RESULT 9

S04137

genome polyprotein - human hepatitis A virus (strain LCD-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.



C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinstone  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905; MUID:85166289  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-854/Product: coat protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKKMGFSAKISLF 20  
|||||  
Db 823 VLPPRRKKMGFSAKISLF 842

RESULT 3  
GNMYM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with di  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M4707; NID:9329582; PIDN:AAA45465.1; PID:9329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKKMGFSAKISLF 20  
|||||  
Db 823 VLPPRRKKMGFSAKISLF 842

RESULT 4  
GNMYR  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro

NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Matjarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MUID:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAJ>  
A:Cross-references: GB:A02990; NID:9329596; PIDN:AAA45472.1; PID:9329597  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKKMGFSAKISLF 20  
|||||  
Db 823 VLPPRRKKMGFSAKISLF 842

RESULT 5  
GNMYK  
genome polyprotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstone, S.M.; Purcell  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison  
A:Reference number: A94149; MUID:87115701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M1632; NID:9329594; PIDN:AAA45471.1; PID:9329595  
A:Note: Submitted to Genbank, August 1987  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:17 ; Search time 108.75 Seconds  
(without alignments)  
17,672 Million cell updates/sec

Title: US-09-171-432a-42

Perfect score: 101

Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : 1: pirl:\*

2: pirl:\*

3: pirl:\*

4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	852	1 GNNYHA	genome polypolyprotein
2	101	100.0	1358	1 A03905	genome polypolyprotein
3	101	100.0	2227	1 GNNYHM	genome polypolyprotein
4	101	100.0	2227	1 GNNYHR	genome polypolyprotein
5	101	100.0	2227	1 GNNYMK	genome polypolyprotein
6	96	95.0	2230	1 GNNYSA	genome polypolyprotein
7	95	94.1	2227	1 GNNYHB	genome polypolyprotein
8	84	83.2	839	1 GNNYS2	genome polypolyprotein
9	75	74.3	341	1 S04137	genome polypolyprotein
10	52.5	52.0	1025	2 A54718	genome polypolyprotein
11	47	46.5	414	2 D66838	diacylglycerol kinase
12	45	44.6	79	2 AG1062	unknown protein T2
13	45	44.6	321	2 E86423	bacteriophage gene
14	44	43.6	340	1 WMBE57	probable 60S ribos
15	44	43.6	343	2 S74937	ribonucleoside-dip
16	44	43.6	369	2 S77299	hypothetical prote
17	44	43.6	775	2 T48957	C4-dicarboxylase-b
18	43	42.6	189	2 A12534	hypothetical prote
19	43	42.6	309	2 D75008	transposase all756
20	43	42.6	379	2 T49919	methanol dehydroge
21	43	42.6	389	2 S41748	hypothetical prote
22	42.5	42.1	105	2 B72598	heat shock protein
23	42	41.6	109	2 S69307	hypothetical prote
24	42	41.6	165	2 S02053	probable membrane
25	42	41.6	178	2 T02335	hypothetical prote
26	42	41.6	834	2 F82673	ribosomal protein
27	41.5	41.1	304	2 H82984	hypothetical prote
28	41	40.6	83	2 T17809	hypothetical prote
29	41	40.6	159	2 C72210	conserved hypothet

30	41	40.6	213	2 H86638	hypothetical prote
31	41	40.6	229	2 T19153	hypothetical prote
32	41	40.6	237	2 C84065	transcription regu
33	41	40.6	370	2 G97001	endoglucanase fam1
34	41	40.6	561	2 G96752	unknown protein F2
35	41	40.6	622	2 T37257	hypothetical prote
36	41	40.6	897	2 T06540	polypolynucleotide
37	41	40.6	1040	2 A34695	axonal glycoprotei
38	41	40.6	1234	2 T30254	hypothetical prote
39	40.5	40.1	924	2 T25007	jumonji protein -
40	40.5	40.1	932	2 T25008	hypothetical prote
41	40	39.6	177	2 C64133	hypothetical prote
42	40	39.6	238	2 C83240	mercuric resistanc
43	40	39.6	259	2 F64532	probable transcrip
44	40	39.6	272	2 T49070	conserved hypothet
45	40	39.6	292	2 S23239	probable heat shoc
46	40	39.6	300	2 AG1885	hypothetical prote
47	40	39.6	310	2 A55053	hypothetical prote
48	40	39.6	310	2 B55053	endothelial monocy
49	40	39.6	371	2 I46089	endothelial monocy
50	40	39.6	372	2 S53724	thyroid transcript
51	40	39.6	372	2 S12002	thyroid-specific
52	40	39.6	380	2 T20269	thyroid nuclear fa
53	40	39.6	401	2 G02321	hypothetical prote
54	40	39.6	448	2 D64367	thyroid transcript
55	40	39.6	523	2 S53945	glutamate dehydrog
56	40	39.6	562	2 H69545	hypothetical prote
57	40	39.6	619	2 T03143	probable fatty-ac1
58	40	39.6	620	2 T50232	R-transactivator p
59	40	39.6	645	2 D98319	actin-like protein
60	40	39.6	645	2 A12963	hypothetical prote

## ALIGNMENTS

RESULT 1  
GNNYHA  
genome polypolyprotein - human hepatitis A virus (strain CR326) (fragment)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904  
R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904; MUID:85185648  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <LIN>  
A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593  
C:Superfamily: hepatitis A virus genome polypolyprotein  
C:Keywords: coat protein; core protein; polypolyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 101; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||  
DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 2  
A03905  
genome polypolyprotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; prote

**THIS PAGE BLANK (USPTO)**



```

; INFORMATION FOR SEQ. ID NO: 25
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 853 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-638-911A-25

```

Query Match	40.68;	Score 41;	DB 1;	Length 853;
Best Local Similarity	46.78;	Pred. No. 69;		
Matches 7;	Conservative	3;	Mismatches 5;	Indels 0;
			Gaps	0

```
QY      3 PPRKMKGLFSQAKI 17
          |||  ||:| :| :
DB     304 PPPVPAKGVFKEAPL 318
```

```

US-08-638-911A-27
US-08-638-911A-27
Sequence 27, Application US/08638911A
Patent No. 5766916
GENERAL INFORMATION:
APPLICANT: Balaevy, Alexander S.
APPLICANT: Chong, Susan M.
TITLE OF INVENTION: Hepatitis G Virus Protease
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/638,911A
FILING DATE: 04/24/96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Evans, Susan T.
REGISTRATION NUMBER: 38,443
REFERENCE/DOCKET NUMBER: 4600-0124
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 853 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-638-911A-27

```

Query Match	40.6%	Score 41	DB 1	Length 853
Best Local Similarity	46.7%	Pred. No. 69		
Matches 7	Conservative 3	Mismatches 5	Indels 0	Gaps 0

```
QY      3 PPRKMGLEFSQAKI 17
          |||  ||:|:|:|
Db     304 PPRVPAKGVEFKEAPL 318
```

RESULT 25  
US-08-638-911A-29  
; Sequence 29, Application US/08638911A  
; Patent No. 5766916  
; GENERAL INFORMATION:

```

1  APPLICANT: Belayev, Alexander S.
2  APPLICANT: Chong, Susan M.
3  TITLE OF INVENTION: Hepatitis G Virus Protease
4  NUMBER OF SEQUENCES: 56
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Dehlinger & Associates
7  STREET: 350 Cambridge Avenue, Suite 250
8  CITY: Palo Alto
9  STATE: CA
10 COUNTRY: USA
11 ZIP: 94306
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.2.2
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/638,911A
19 FILING DATE: 04/24/96
20 CLASSIFICATION: 435
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Evans, Susan T.
23 REGISTRATION NUMBER: 38,443
24 REFERENCE/DOCKET NUMBER: 4600-0124
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (415) 324-0880
27 TELEFAX: (415) 324-0960
28 INFORMATION FOR SEQ ID NO: 29:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 853 amino acids
31 TYPE: amino acid
32 TOPOLOGY: linear
33 MOLECULE TYPE: protein
34 US-08-638-911A-29

```

Query Match	40.68	Score 41	DB 1	Length 853
Best Local Similarity	46.78	Pred. No. 69		
Matches 7	Conservative 3	Mismatches 5	Indels 0	Gaps 0

```
QY      3 PPPRKMGLFSQAKI 17
          ||| ||:| :| :
Db     304 PPPVPAKGVFKEAPL 318
```

Search completed: June 16, 2002, 00:03:16  
Job time: 8530 sec

GENERAL INFORMATION:  
APPLICANT: TITLE OF INVENTION: Detection of Viral Antigens Coded  
TITLE OF INVENTION: by Reverse Reading Frames  
NUMBER OF SEQUENCES: 157  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06266  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,561  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0202.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-06266-89

Query Match 40.6%; Score 41; DB 5; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKI 17  
Db 57 PPPVPAKGVFKKAPL 71

RESULT 22  
US-08-638-911A-35  
Sequence 35, Application US/08638911A  
Patent No. 5766916  
GENERAL INFORMATION:  
APPLICANT: Belaeyv, Alexander S.  
APPLICANT: Chong, Susan M.  
TITLE OF INVENTION: Hepatitis G Virus Protease  
NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/638,911A  
FILING DATE: 04/24/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4600-0124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 616 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-638-911A-35

Query Match 40.6%; Score 41; DB 1; Length 616;  
Best Local Similarity 46.7%; Pred. No. 49;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKI 17  
Db 184 PPPVPAKGVFKKAPL 198

RESULT 23  
US-08-638-911A-25  
Sequence 25, Application US/08638911A  
Patent No. 5766916  
GENERAL INFORMATION:  
APPLICANT: Belaeyv, Alexander S.  
APPLICANT: Chong, Susan M.  
TITLE OF INVENTION: Hepatitis G Virus Protease  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/638,911A  
FILING DATE: 04/24/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4600-0124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960

APPLICANT: Young, Lavonne M.  
APPLICANT: Fry, Kirk E.  
APPLICANT: Linen, Jeffrey M.  
TITLE OF INVENTION: Hepatitis G Virus and Molecular  
TITLE OF INVENTION: Cloning Thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,361  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,543  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-461-361-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPPRKMGLFSQAKI 17  
||| ||:| :| :  
Db 57 PVPVPAKGVFKKAPL 71

RESULT 20  
US-08-485-910-106  
; Sequence 106, Application US/08485910  
; Patent No. 5874563  
; GENERAL INFORMATION:

APPLICANT: Kim, Jungshuh P.  
APPLICANT: Wages, John  
APPLICANT: Young, Lavonne M.  
APPLICANT: Fry, Kirk E.  
APPLICANT: Linen, Jeffrey M.  
TITLE OF INVENTION: Hepatitis G Virus and Molecular  
TITLE OF INVENTION: Cloning Thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,910  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,543  
FILING DATE: 03-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,985  
FILING DATE: 20-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-910-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPPRKMGLFSQAKI 17  
||| ||:| :| :  
Db 57 PVPVPAKGVFKKAPL 71

RESULT 21  
PCT-US95-06266-89  
; Sequence 89, Application PC/TUS9506266

;; TITLE OF INVENTION: Cloning Thereof  
;; NUMBER OF SEQUENCES: 277  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dehlinger & Associates  
;; STREET: 350 Cambridge Ave., Suite 250  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94306  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/444,733  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/389,886  
;; FILING DATE: 15-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/357,509  
;; FILING DATE: 16-DEC-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,729  
;; FILING DATE: 26-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/344,271  
;; FILING DATE: 23-NOV-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,558  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,543  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/246,985  
;; FILING DATE: 20-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 106:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 108 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-444-733-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPPRKMKGLEFSOAKI 17  
||| ||:| :| :  
Db 57 PPPVPAKGVFKEAPL 71

RESULT 18  
US-08-464-134-106  
; Sequence 106, Application US/08464134  
; Patent No. 3849532  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Jungsuh P.  
; APPLICANT: Wages, John  
; APPLICANT: Young, Lavonne M.  
; APPLICANT: Fry, Kirk E.

;; APPLICANT: Linnen, Jeffrey M.  
;; TITLE OF INVENTION: Hepatitis G Virus and Molecular  
;; Cloning Thereof  
;; NUMBER OF SEQUENCES: 277  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dehlinger & Associates  
;; STREET: 350 Cambridge Ave., Suite 250  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94306  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/464,134  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/389,886  
;; FILING DATE: 15-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/357,509  
;; FILING DATE: 16-DEC-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,729  
;; FILING DATE: 26-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/344,271  
;; FILING DATE: 23-NOV-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,558  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,543  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/246,985  
;; FILING DATE: 20-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 106:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 108 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-464-134-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPPRKMKGLEFSOAKI 17  
||| ||:| :| :  
Db 57 PPPVPAKGVFKEAPL 71

RESULT 19  
US-08-461-361-106  
; Sequence 106, Application US/08461361  
; Patent No. 5856134  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Jungsuh P.  
; APPLICANT: Wages, John  
; APPLICANT: Young, Lavonne M.  
; APPLICANT: Fry, Kirk E.



OY 2 LPPPKMKGLF 12  
||| | : | : |  
Db 4 LPPPKMLPGIF 14

## RESULT 15

PCT-US91-03388-4  
; Sequence 4, Application PC/TUS9103388  
; GENERAL INFORMATION:  
; APPLICANT: Hewick, Rodney M.  
; APPLICANT: Wang, Jack H.  
; TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/03388  
; FILING DATE: 19910515  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kapiros, Ellen J.  
; REGISTRATION NUMBER: 32,245  
; REFERENCE/DOCKET NUMBER: G15182X-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-876-1170  
; TELEFAX: 617-876-5851  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bos taurus  
; TISSUE TYPE: Bone  
; PCT-US91-03388-4

Query Match 44.6%; Score 45; DB 5; Length 23;  
Best Local Similarity 63.6%; Pred. No. 0.29;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 2 LPPPKMKGLF 12  
||| | : | : |  
Db 4 LPPPKMLPGIF 14

## RESULT 16

US-08-466-033-106  
; Sequence 106, Application US/08466033  
; Patent No. 5766840  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Jungsuh P.  
; APPLICANT: Wages, John  
; APPLICANT: Young, Lavonne M.  
; APPLICANT: Fry, Kirk E.  
; APPLICANT: Linnen, Jeffrey M.  
; TITLE OF INVENTION: Hepatitis G Virus and Molecular  
; Cloning Thereof  
; NUMBER OF SEQUENCES: 277

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,033  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/389,886  
; FILING DATE: 15-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/357,509  
; FILING DATE: 16-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,729  
; FILING DATE: 26-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,271  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/285,558  
; FILING DATE: 03-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/285,543  
; FILING DATE: 03-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/246,985  
; FILING DATE: 20-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 106:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 108 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-466-033-106

Query Match 40.6%; Score 41; DB 1; Length 108;  
Best Local Similarity 46.7%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
OY 3 PPKMKGLFSQAKI 17  
||| | : | : |  
Db 57 PPKMKGLFSQAKI 71

## RESULT 17

US-08-444-733-106  
; Sequence 106, Application US/08444733  
; Patent No. 5824507  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Jungsuh P.  
; APPLICANT: Wages, John  
; APPLICANT: Young, Lavonne M.  
; APPLICANT: Fry, Kirk E.  
; APPLICANT: Linnen, Jeffrey M.  
; TITLE OF INVENTION: Hepatitis G Virus and Molecular

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO.: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-4

Query Match 52.0%; Score 52.5; DB 5; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.97;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPRKMGKLSOAKISLF 20  
DB 176 LPPPRKMGKLSOAKISLF 193

RESULT 13  
PCT-US95-04567-2  
Sequence 2, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROXYIMIDINE DEHYDROGENASE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO.: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-2

Query Match 45.0%; Score 45.5; DB 5; Length 1025;  
Best Local Similarity 61.1%; Pred. No. 15;  
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPRKMGKLSOAKISLF 19  
DB 176 LPPPRKMGKLSOAKISLF 192

RESULT 14  
US-07-800-364B-4  
Sequence 4, Application US/07800364B  
Patent No. 5688678  
GENERAL INFORMATION:  
APPLICANT: Hewick, Rodney M.  
APPLICANT: Wang, Jack H.  
APPLICANT: Wozney, John M.  
TITLE OF INVENTION: Celeste, Anthony J.  
TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/800,364B  
FILING DATE: 26-NOV-1991  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kapinos, Ellen J.  
REGISTRATION NUMBER: 32,245  
REFERENCE/DOCKET NUMBER: GI 5182A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-876-1170  
TELEFAX: 617-876-5851  
INFORMATION FOR SEQ ID NO.: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus  
TISSUE TYPE: Bone  
US-07-800-364B-4

Query Match 44.6%; Score 45; DB 1; Length 23;  
Best Local Similarity 63.6%; Pred. No. 0.29;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

APPLICANT: FERNANDEZ-SALGUERO, Pedro  
TITLE OF INVENTION: CLONING AND EXPRESSION OF cDNA FOR HUMAN  
DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,309  
FILING DATE: 09-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 15280-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-304-309-2

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.97;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPKMKGLFSQAKISLF 20  
||||| : |||||  
Db 176 LPPEKMSKSEAYS-AKIALF 193

RESULT 10  
US-08-991-942-2  
Sequence 2, Application US/08991942  
Patent No. 6015673  
GENERAL INFORMATION:  
APPLICANT: GONZALEZ, Frank J.  
APPLICANT: FERNANDEZ-SALGUERO, Pedro  
TITLE OF INVENTION: CLONING AND EXPRESSION OF cDNA FOR HUMAN  
DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,942  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,309

FILING DATE: 09-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 15280-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-942-2

Query Match 52.0%; Score 52.5; DB 3; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.97;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPKMKGLFSQAKISLF 20  
||||| : |||||  
Db 176 LPPEKMSKSEAYS-AKIALF 193

RESULT 11  
US-09-138-103-2  
Sequence 2, Application US/09138103A  
Patent No. 6232448  
GENERAL INFORMATION:  
APPLICANT: Yoshikubo, Takashi  
APPLICANT: Hasegawa, Masami  
TITLE OF INVENTION: Immunological Materials and Methods for Detecting  
DIHYDROPYRIMIDINE DEHYDROGENASE  
FILE REFERENCE: 09/138,103 Yoshikubo, et al.  
CURRENT APPLICATION NUMBER: US/09/138,103A  
CURRENT FILING DATE: 1998-08-21  
EARLIER APPLICATION NUMBER: 97114630.3  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1025  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-138-103-2

Query Match 52.0%; Score 52.5; DB 4; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.97;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPKMKGLFSQAKISLF 20  
||||| : |||||  
Db 176 LPPEKMSKSEAYS-AKIALF 193

RESULT 12  
PCT-US95-04567-4  
Sequence 4, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210

TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US1  
CURRENT APPLICATION NUMBER: US/08/397,232A  
CURRENT FILING DATE: 1995-04-17  
EARLIER APPLICATION NUMBER: 07/947,338  
EARLIER FILING DATE: 1992-09-18  
EARLIER APPLICATION NUMBER: PCT/US93/08610  
EARLIER FILING DATE: 1993-09-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO: 4  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRKMKGFSQAKISLF 20  
|||||  
Db 823 VLPPRKMKGFSQAKISLF 842

RESULT 7  
US-09-171-387-2  
Sequence 2, Application US/09171387  
Patent No. 6280734  
GENERAL INFORMATION:  
APPLICANT: RAYCHAUDHURI, GOPA;  
EMERSON, SUZANNE, U.;  
PURCELL, ROBERT, H.  
TITLE OF INVENTION: SIMIAN-HUMAN HAV  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,387  
FILING DATE: 24-Mar-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/06506  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US60/015,642  
FILING DATE: 19-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feiler  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-42290S1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2227 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRKMKGFSQAKISLF 20  
|||||  
Db 823 VLPPRKMKGFSQAKISLF 842

RESULT 8  
US-08-087-016-2  
Sequence 2, Application US/08087016  
Patent No. 5430135  
GENERAL INFORMATION:  
APPLICANT: NAINAN, OMANA V.  
APPLICANT: MARGOLIS, HAROLD S.  
APPLICANT: ROBERTSON, BETTY H.  
APPLICANT: BRINTON, MARGO H.  
APPLICANT: EBERT, JAMES W.  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L Street N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,016  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,828  
FILING DATE: 03-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 839 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 3.8e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRKMKGFSQAKIS 18  
|||||  
Db 822 ILPPRKLKGLFSQSKIS 839

RESULT 9  
US-08-304-309-2  
Sequence 2, Application US/08304309  
Patent No. 5856454  
GENERAL INFORMATION:  
APPLICANT: CONZALEZ, Frank J.

```
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2
```

```
Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VLPPRRKKMKGLEFSQAKISLF 20
    ||||||||||||||||
Db 823 VLPPRRKKMKGLEFSQAKISLF 842
```

```
RESULT 3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4
```

```
Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VLPPRRKKMKGLEFSQAKISLF 20
    ||||||||||||||||
Db 823 VLPPRRKKMKGLEFSQAKISLF 842
```

```
RESULT 4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
```

```
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6
```

```
Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VLPPRRKKMKGLEFSQAKISLF 20
    ||||||||||||||||
Db 823 VLPPRRKKMKGLEFSQAKISLF 842
```

```
RESULT 5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2
```

```
Query Match          100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VLPPRRKKMKGLEFSQAKISLF 20
    ||||||||||||||||
Db 823 VLPPRRKKMKGLEFSQAKISLF 842
```

```
RESULT 6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:15 ; Search time 79.04 Seconds

(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-42

Perfect score: 101

Sequence: 1 VLPPRRMKGLFSQAKISLP 20

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 60 summaries

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	1091	6	Patent No. 5516630-2
2	101	100.0	2227	3	US-08-475-886-2
3	101	100.0	2227	3	US-08-475-886-4
4	101	100.0	2227	3	US-08-475-886-6
5	101	100.0	2227	4	US-08-397-232-2
6	101	100.0	2227	4	US-08-397-232-4
7	101	100.0	2227	4	US-09-171-387-2
8	84	83.2	839	1	US-08-087-016-2
9	52.5	52.0	1025	2	US-08-304-309-2
10	52.5	52.0	1025	3	US-08-991-942-2
11	52.5	52.0	1025	4	US-09-138-103-2
12	52.5	52.0	1025	5	PCR-US95-04567-4
13	45.5	45.0	1025	5	PCR-US95-04567-2
14	45.5	45.0	23	1	US-07-800-364B-4
15	45	44.6	23	5	PCR-US91-03388-4
16	41	40.6	108	1	US-08-466-033-106
17	41	40.6	108	2	US-08-444-733-106
18	41	40.6	108	2	US-08-464-134-106
19	41	40.6	108	2	US-08-461-361-106
20	41	40.6	108	2	US-08-485-910-106
21	41	40.6	108	5	PCR-US95-06266-89
22	41	40.6	616	1	US-08-638-911A-35
23	41	40.6	853	1	US-08-638-911A-25
24	41	40.6	853	1	US-08-638-911A-27
25	41	40.6	853	1	US-08-638-911A-29
26	41	40.6	853	1	US-08-638-911A-31
27	41	40.6	2873	1	US-08-466-033-15

28	41	40.6	2873	1	US-08-638-911A-2	Sequence 2, Appl
29	41	40.6	2873	2	US-08-444-733-15	Sequence 15, Appl
30	41	40.6	2873	2	US-08-464-134-15	Sequence 15, Appl
31	41	40.6	2873	2	US-08-461-361-15	Sequence 15, Appl
32	41	40.6	2873	2	US-08-485-910-15	Sequence 15, Appl
33	41	40.6	2873	5	PCR-US95-06266-15	Sequence 15, Appl
34	41	40.6	2910	1	US-08-466-033-183	Sequence 183, App
35	41	40.6	2910	2	US-08-444-733-183	Sequence 183, App
36	41	40.6	2910	2	US-08-464-134-183	Sequence 183, App
37	41	40.6	2910	2	US-08-461-361-183	Sequence 183, App
38	41	40.6	2910	2	US-08-485-910-183	Sequence 183, App
39	41	40.6	2910	5	PCR-US95-06266-157	Sequence 157, App
40	40	39.6	34	2	US-08-751-767A-38	Sequence 38, Appl
41	40	39.6	166	1	US-08-129-456A-11	Sequence 11, Appl
42	40	39.6	166	4	US-08-360-821B-11	Sequence 11, Appl
43	40	39.6	310	1	US-08-129-456A-36	Sequence 36, Appl
44	40	39.6	310	1	US-08-129-456A-37	Sequence 37, Appl
45	40	39.6	310	2	US-08-705-868-3	Sequence 3, Appl
46	40	39.6	310	3	US-09-123-615-3	Sequence 3, Appl
47	40	39.6	310	4	US-08-360-821B-35	Sequence 35, Appl
48	40	39.6	312	4	US-08-360-821B-36	Sequence 36, Appl
49	40	39.6	371	2	US-08-442-809A-76	Sequence 76, Appl
50	40	39.6	391	2	US-08-644-034A-1	Sequence 1, Appl
51	39	38.6	100	4	US-08-818-112-67	Sequence 67, Appl
52	39	38.6	100	4	US-08-818-111-68	Sequence 68, Appl
53	39	38.6	100	4	US-09-056-556-67	Sequence 67, Appl
54	39	38.6	347	4	US-09-094-557-3	Sequence 3, Appl
55	39	38.6	607	2	US-08-878-989-15	Sequence 15, Appl
56	39	38.6	607	4	US-09-272-796-15	Sequence 15, Appl
57	39	38.6	1248	3	US-08-726-214-16	Sequence 16, Appl
58	38	37.6	205	4	US-09-385-259-3	Sequence 3, Appl
59	38	37.6	229	1	US-08-287-959-7	Sequence 7, Appl
60	38	37.6	442	3	US-08-834-306-52	Sequence 52, Appl

## ALIGNMENTS

RESULT 1  
5516630-2  
Patent No. 5516630  
APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; PEINSTONE,  
STEPHEN M.; PURCELL, ROBERT H.; RACANIELLO, VINCENT R.;  
BARODDY, BAIGIE M.  
TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/788,262  
FILING DATE: 06-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 256,135  
FILING DATE: 06-OCT-1988  
APPLICATION NUMBER: 654,942  
FILING DATE: 27-SEP-1984  
APPLICATION NUMBER: 537,911  
FILING DATE: 30-SEP-1983  
SEQ ID NO: 2:  
LENGTH: 1091  
5516630-2

Query Match 100.0%; Score 101; DB 6; length 1091;  
Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPRRMKGLFSQAKISLP 20  
DB 1060 VLPPRRMKGLFSQAKISLP 1079

RESULT 2  
US-08-475-886-2  
; sequence 2, Application US/08475886A

Query Match 44.68; Score 45; DB 12; Length 23;  
Best Local Similarity 63.68; Pred. No. 2.4;  
Matches 7; Conservative 2; Mismatches 0; Indels 0;  
Gaps 0;  
QY 2 LPPRRKKGLF 12  
|||||:|:|  
Db 4 lppnklpgif 14

Search completed: June 16, 2002, 00:01:44  
Job time: 12733 sec

CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 114 AA;

Query Match 45.0%; Score 45.5; DB 22; Length 114;  
 Best Local Similarity 44.4%; Pred. NO. 9.9;  
 Matches 8; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 3 PPPRRKGLFSQAKISLF 20  
 ||| | | | : : : :  
 DB 29 PPPK-kgfikspvgvf 45

RESULT 24  
 AAR86000  
 ID AAR86000 standard; Protein; 1025 AA.  
 AC AAR86000;  
 XX  
 DT 04-JUN-1996 (first entry)  
 XX  
 DE Bovine dihydropyrimidine dehydrogenase.  
 XX  
 KW Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;  
 KW pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;  
 KW fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;  
 KW frameshift mutation.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key. Location/Qualifiers  
 FT Peptide 788..795  
 FT /note= "Shows similarity to flavin binding domain  
 FT of dihydrocrotonate"  
 FT Peptide 187..204  
 FT /note= "Shows similarity to flavin-NADPH binding  
 FT domains of thioredoxin reductase"  
 FT Peptide 332..348  
 FT /note= "Shows similarity to flavin-NADPH binding  
 FT domains of thioredoxin reductase"  
 XX  
 PN W09528489-A1.  
 XX  
 PD 26-OCT-1995.  
 XX  
 PF 13-APR-1995; 95WO-US04567.  
 XX  
 PR 13-APR-1994; 94US-0227357.  
 XX  
 PA (UABR-) UAB RES FOUND.  
 XX  
 PI Cheng X, Diasio RB, Johnson M, Lu Z, Zhang R;  
 XX  
 DR WPI: 1995-373803/48.  
 DR N-PSDB: AAT03132 and AAT03142.  
 XX  
 PT Novel dihydropyrimidine dehydrogenase gene - used to optimise  
 PT 5-fluorouracil doses given to cancer patients  
 XX  
 PS Claim 3; Page 115-32; 207pp; English.  
 XX  
 CC This sequence represents bovine liver dihydropyrimidine dehydrogenase  
 CC (DPD). DPD catalyses the initial and rate limiting step in pyrimidine  
 CC catabolism, the reduction of pyrimidines to 5,6-dihydropyrimidines.  
 CC DPD is a complex enzyme consisting of two identical subunits, containing  
 CC FMN, FAD and iron-sulphur centers, and utilising NADPH as a cofactor.

CC DPD has also been shown to catalyse the reduction of various pyrimidine  
 CC analogues including the fluoropyrimidine anticancer drug 5-fluorouracil  
 CC (Fura). Up to 85% of administered Fura may be catabolised by DPD, and  
 CC it therefore governs the effectiveness of Fura as an anticancer drug.  
 CC DPD genes or fragments of them may be used in the detection of DPD in a  
 CC sample, esp. isolated from a cancer patient. According to the amount  
 CC of DPD detected, a therapeutically effective amount of Fura may be  
 CC determined and administered. DPD deficiency, leading to life-  
 CC threatening toxicity on exposure to Fura, in a human caused by a  
 CC frameshift mutation may be determined by means of a molecular biological  
 CC assay to detect the deletion of an A residue at codon 318 within the  
 CC DPD-coding region.  
 XX  
 SQ Sequence 1025 AA;

Query Match 45.0%; Score 45.5; DB 16; Length 1025;  
 Best Local Similarity 61.1%; Pred. NO. 90;  
 Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPRKMKGLFSQAKISL 19  
 ||| | | | : | | | :  
 DB 176 lpppekmpays-akial 192

RESULT 25  
 AAR15517  
 ID AAR15517 standard; Protein; 23 AA.  
 AC AAR15517;  
 XX  
 DT 09-MAR-1992 (first entry)  
 XX  
 DE BMP-8 peptide.  
 XX  
 KW Cartilage; wound healing; tissue repair; BMP.  
 XX  
 OS Bos taurus.  
 XX  
 PN W09118098-A.  
 XX  
 PD 28-NOV-1991.  
 XX  
 PF 15-MAY-1991; 91WO-US03388.  
 XX  
 PR 15-JAN-1991; 91US-0641204.  
 PR 16-MAY-1990; 90US-0525357.  
 XX  
 PA (GENE-) GENETICS INST INC.  
 XX  
 PI Hewick RM, Wang JH;  
 XX  
 DR WPI: 1991-369252/50  
 DR N-PSDB: AAO15240, AAO15242, AAO15245, AAO15246.  
 XX  
 PT New BMP-8 protein - useful in inducing cartilage and/or bone  
 PT formation to treat wounds and repair fractures and tissues, e.g.  
 PT burns, incisions and ulcers  
 XX  
 PS Claim 1(d); Page 46; 50pp; English.  
 XX  
 CC This sequence shares some homology (i.e. Asn-Glu-Leu-Pro)  
 CC with BMP-3 (see W088/00205 and W089/10409).  
 CC Pharmaceutical comps. contg. BMP-8, which comprises at least one  
 CC of the fragments represented in AAR15517 and AAR15522, can be used to  
 CC aid bone and/or cartilage formation or wound healing and tissue repair.  
 CC The proteins are not very species specific so can be used in domestic  
 CC and farm animals as well as humans.  
 CC See also AAO15240-48, AAR15517 and AAR15522.  
 XX  
 SQ Sequence 23 AA;



PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 46.5%; Score 47; DB 21; Length 407;  
Best Local Similarity 56.2%; Pred. No. 20;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PPRKMKGLFSQAKIS 18  
||| | | | | | | | | |  
Db 99 PPRdmgdlkxvnl 114

## RESULT 22

AA000818  
ID AA000818 standard; Protein: 117 AA.

XX AC AA000818;

DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 14710.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN W0200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AAI80749.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 14710; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SO Sequence 117 AA;

Query Match 45.5%; Score 46; DB 22; Length 117;  
Best Local Similarity 52.9%; Pred. No. 8.4;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 3 PPRKMKGLFSQAKISL 19  
||| | | | | | | | | |  
Db 24 PPRKMKGLFSQAKISL 40

## RESULT 23

AA003241  
ID AA003241 standard; Protein: 114 AA.

XX AC AA003241;

DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 17133.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN W0200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AAI83172.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 17133; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132864.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0133124.  
PR 21-MAY-1999; 99US-0133353.  
PR 24-MAY-1999; 99US-0133629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138099.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144003.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145986.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147039.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154409.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Binding-site /label= NADP(H)\_binding site  
 FT Binding-site 335..351  
 FT Binding-site 472..482  
 FT Binding-site /label= FAD\_binding\_site  
 FT Binding-site 661..678  
 FT Binding-site /label= Uracil\_binding\_site  
 FT Binding-site 953..964  
 FT Binding-site /label= 4Fe-4S\_binding\_site  
 FT Binding-site 986..997  
 FT Binding-site /label= 4Fe-4S\_binding\_site  
 XX  
 PN MO608568-A2.  
 XX  
 PD 21-MAR-1996.  
 XX  
 PF 07-SEP-1995; 95WO-US12016.  
 XX  
 PR 12-SEP-1994; 94US-0304309.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Fernandez-Salguero P, Gonzalez FJ;  
 XX  
 DR WPI: 1996-179943/18.  
 DR N-PSDB; AAT14077.  
 XX  
 PT Pig and human dihydro:pyrimidine dehydrogenase (DPD) genes and  
 PT probes - useful for detection of DPD deficiencies and  
 PT identification of humans at risk of toxic reaction to  
 PT 5-fluoro:uracil anti-cancer treatment  
 XX  
 PS Claim 7; Page 48-51; 78pp; English.  
 XX  
 CC Human dihydro:pyrimidine dehydrogenase (DPD) (AAR91420) can be  
 CC expressed as a recombinant protein in a prokaryotic host cell  
 CC by insertion of an encoding cDNA sequence (see AAT14077) into a  
 CC vector, transfection of host cells, and culturing the cells in  
 CC medium contg. uracil, FAD and FMN. DPD is used to raise  
 CC antibodies useful in competitive binding immunoassays.  
 CC  
 XX  
 SQ Sequence 1025 AA;  
 XX

Query Match 52.0%; Score 52.5; DB 17; Length 1025;  
 Best Local Similarity 63.2%; Pred. No. 6.9;  
 Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPRKMGLEFSQAKISLF 20  
 DB 176 lpppkmseays-aktalf 193  
 ||||| :| |||:| |  
 ||||| :| |||:| |

RESULT 20  
 AAW93361  
 ID AAW93361 standard; Protein: 1025 AA.  
 XX  
 AC AAW93361;  
 XX  
 DT 28-MAY-1999 (first entry)  
 XX  
 DE Human DPD protein.  
 XX  
 KW DPD; dihydro:pyrimidine-dehydrogenase; monoclonal antibody; Mab;  
 KW immunoassay reagent; cancer patient; treatment; antitumor agent;  
 KW 5-fluorouracil; affinity purification; toxicity.  
 OS Homo sapiens.  
 XX  
 PN DE19837391-A1.  
 XX

PD 25-FEB-1999.  
 XX  
 PF 18-AUG-1998; 98DE-1037391.  
 XX  
 PR 22-AUG-1997; 97EP-0114630.  
 XX  
 PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Hasegawa M, Yoshikubo T;  
 XX  
 DR WPI: 1999-155202/14.  
 DR N-PSDB; AAX22903.  
 XX  
 PT Monoclonal antibody specific for dihydro:pyrimidine dehydrogenase -  
 PT for assessing patient response to 5-fluorouracil antitumor agents  
 XX  
 PS Disclosure; Page 17-22; 34pp; German.  
 XX  
 CC This invention describes a monoclonal antibody (Mab) specific for  
 CC dihydro:pyrimidine dehydrogenase (DPD). This Mab is used as immunoassay  
 CC reagents to identify a lack of DPD in a patient and to assess the  
 CC sensitivity of cancer patients to treatment with antitumor agents of the  
 CC 5-fluorouracil (5-FU) type. The Mab can also be used for affinity  
 CC purification of DPD. DPD is involved in reduction of 5-FU (and related  
 CC catabolites and derivatives) and lack of it is associated with increased  
 CC toxicity of this type of antitumor agent. It has specific binding  
 CC interaction. The Mab provide a sensitive and reliable test for DPD,  
 CC which is simple, rapid and suitable for routine screening.  
 CC  
 XX  
 SQ Sequence 1025 AA;  
 XX

Query Match 52.0%; Score 52.5; DB 20; Length 1025;  
 Best Local Similarity 63.2%; Pred. No. 6.9;  
 Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 2 LPPPRKMGLEFSQAKISLF 20  
 DB 176 lpppkmseays-aktalf 193  
 ||||| :| |||:| |  
 ||||| :| |||:| |

RESULT 21  
 AAG06085  
 ID AAG06085 standard; Protein: 407 AA.  
 XX  
 AC AAG06085;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 2730.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 XX

RESULT 17  
AAB69447  
ID AAB69447 standard; Peptide; 26 AA.  
XX  
AC AAB69447;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Synthetic HAV P2A peptide, SEQ ID NO: 47.  
XX  
KM Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XX antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
XX Synthetic.  
XX  
PN WO200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000WO-US19267.  
XX  
PR 15-JUL-1999; 99US-0144412.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI: 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
XX detecting anti-hepatitis A virus and as vaccines -  
XX  
PS Claim 13; Page 98; 130pp; English.  
XX  
CC The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
XX  
SQ Sequence 26 AA:

Query Match 64.4%; Score 65; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPPPRKMGKGLF 12  
DB 14 VLPPTKMGKGLF 25

RESULT 18  
AAR86001  
ID AAR86001 standard; Protein; 1025 AA.  
XX  
AC AAR86001;  
XX  
DT 04-JUN-1996 (first entry)

XX  
DE Human dihydropyrimidine dehydrogenase.  
XX  
KM Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;  
KM pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;  
KM fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;  
KM frameshift mutation.  
XX  
OS Homo sapiens.  
XX  
PN WO9528489-A1.  
XX  
PD 26-OCT-1995.  
XX  
PF 13-APR-1995; 95WO-US04567.  
XX  
PR 13-APR-1994; 94US-0227357.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Cheng X, Diasio RB, Johnson M, Lu Z, Zhang R;  
XX  
DR WPI: 1995-373803/48.  
XX  
PT N-PSDB: AAT03133.  
XX  
PT Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise  
XX 5-fluoro:uracil doses given to cancer patients  
XX  
PS Claim 9; Page 140-56; 207pp; English.  
XX  
CC This sequence represents human liver dihydropyrimidine dehydrogenase  
CC (DPD). DPD catalyses the initial and rate limiting step in pyrimidine  
CC catabolism, the reduction of pyrimidines to 5,6-dihydropyrimidines.  
CC DPD is a complex enzyme consisting of two identical subunits, containing  
CC FMN, FAD and iron-sulphur centers, and utilising NADPH as a cofactor.  
CC DPD has also been shown to catalyse the reduction of various pyrimidine  
CC analogues including the fluoropyrimidine anticancer drug 5-fluorouracil  
CC (Fura). Up to 85% of administered Fura may be catabolised by DPD, and  
CC it therefore governs the effectiveness of Fura as an anticancer drug.  
CC DPD genes or fragments of them may be used in the detection of DPD in a  
CC sample, esp. isolated from a cancer patient. According to the amount  
CC of DPD detected, a therapeutically effective amount of Fura may be  
CC determined and administered. DPD deficiency, leading to life-  
CC threatening toxicity on exposure to Fura, in a human caused by a  
CC frameshift mutation may be determined by means of a molecular biological  
CC assay to detect the deletion of an A residue at codon 318 within the  
CC DPD-coding region.  
XX  
SQ Sequence 1025 AA:

Query Match 52.0%; Score 52.5; DB 16; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 6.9;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMGKGLFSQAKISLF 20  
DB 176 LPPPEKMGSEAY-arkia1f 193

RESULT 19  
AAR91420  
ID AAR91420 standard; Protein; 1025 AA.  
XX  
AC AAR91420;  
XX  
DT 11-JUN-1996 (first entry)  
XX  
DE Human dihydropyrimidine dehydrogenase.  
XX  
KM Dihydropyrimidine dehydrogenase; DPD; DPD gene; probe; deficiency;  
KM 5-fluorouracil; cytostatic; cancer; antibody.  
XX

Best Local Similarity 95.0%; Pred. No. 1.8e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
||||||| |||||||||  
DB 823 vlppprkmgkglfsqakislif 842

RESULT 15  
AAR15629  
ID AAR15629 standard; Protein; 839 AA.  
XX  
AC AAR15629;  
XX  
DT 17-DEC-2001 (updated)  
DT 17-MAR-1992 (first entry)  
XX  
DE Capsid region of cyno-HAV isolate CY-145.  
XX  
KM Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.  
XX  
OS Cynomolgus monkey hepatitis A virus, isolate CY-145.  
XX  
FH Key Location/Qualifiers  
FT Protein 1..245  
FT /label= VP0 246..491  
FT /label= VP3 492..791  
FT /label= VP1 792  
FT /label= P2 /note= "incomplete"  
FT Cleavage-site 245..246  
FT Cleavage-site 491..492  
FT Cleavage-site 791..792  
FT Active-site 315  
FT Active-site 593  
XX  
XX USN7678828-N.  
XX  
PD 12-NOV-1991.  
XX  
PF 03-APR-1991; 91US-0678828.  
XX  
PR 03-APR-1991; 91US-0678828.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN.  
XX  
PI Halnan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;  
DR WPI; 1991-376737/51.  
DR N-PSDB: AAQ15180.  
XX  
XX Hepatitis A virus isolates and DNA - used to prepare vaccines for  
PT preventing hepatitis A virus infection.  
XX  
PS Disclosure: Fig 3; 23pp; English.  
XX  
CC The sequence was deduced from the nucleotide sequence obtd. by PCR  
CC amplification of cyno-HAV viral RNA obtd. from the stool of a  
CC cynomolgus monkey with serologically and histologically confirmed  
CC spontaneous hepatitis A. The sequence differs from the human HAV  
CC isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.  
CC USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val  
CC pair at the VP3-VP1 cleavage site in the human isolate is replaced  
CC by a Gln-Thr pair in the cyno-HAV. The other two cleavage sites are  
CC the same. Two residues have been identified as part of the immuno-  
CC dominant region (see feature table) and are different to those in  
CC the same position in human HAV. The protein and peptides derived  
CC from it can be used in the prepn. of vaccines for the prevention of  
CC HAV infection.  
CC See also AAR15056.

CC (Note: Revised entry submitted to correct the patent number format of  
CC US Government-owned NRTS applications to prevent clashes with ongoing US  
CC granted patent numbers. For further information please visit the Derwent  
CC web site at [www.derwent.com/dwpl/updates/nrtis\\_us.html](http://www.derwent.com/dwpl/updates/nrtis_us.html).)  
XX

SQ Sequence 839 AA;

Query Match 83.2%; Score 84; DB 12; Length 839;  
Best Local Similarity 83.3%; Pred. No. 5.4e-05;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18  
:|||||:|||||:|||  
DB 822 vlppprkikglfsqskis 839

RESULT 16

AAW42930  
ID AAW42930 standard; peptide; 25 AA.

XX  
AC AAW42930;

DT 28-APR-1998 (first entry)

XX Immunogenic Hepatitis A virus peptide YK-1665.

XX Immunogenic peptide; immunogenic epitope; P2A protein;

KW immune response; antibody.

XX Synthetic.

OS Hepatitis A virus.

XX W09740147-A1.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-US06891.

XX 19-APR-1996; 96US-0015644.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

DR WPI; 1997-535831/49.

XX Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an

PT immune response to HAV in a mammal or to detect the presence of

XX antibodies against HAV in a mammal

PS Claim 18; Page 112; 140pp; English.

XX Peptides AAW42922-30 are immunogenic peptides corresponding to

CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are

CC substantially similar to a portion of the amino acid sequence of the P2A

CC protein of HAV corresponding to amino acids 792-980. Compositions

CC containing the peptides can be used to induce an immune response to HAV

CC in a mammal. The peptides can also be used to detect the presence of

CC antibodies against HAV in mammalian serum. The peptides can also be used

CC to make an antibody against HAV by administering the peptide to a

CC mammal.

SQ Sequence 25 AA;

Query Match 64.4%; Score 65; DB 18; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLF 12  
|||||||  
DB 14 vlppprkmgkglf 25

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.

SO Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGFSQAKISLF 20  
DB 823 VLPPPRKMGFSQAKISLF 842

RESULT 13

AAB18609  
ID AAB18609 standard; Protein; 2227 AA.

AC AAB18609;

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
HAV 4380.

OS Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PF 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93MO-US08610.

PR 10-MAR-1995; 95US-0397232.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI: 2000-586464/55.

DR N-PSDB; AAA73478.

PS Disclosure: Columns 93-104; 72pp; English.

The present sequence is derived from a live attenuated hepatitis A  
virus (HAV) of the invention, designated HAV 4380. The sequence is  
produced by modifying wild type HAV strain HM-174. The HAV of the  
invention are adapted to growth in the human fibroblast-like cell  
line MRC-5. The HAV is able to propagate in MRC-5 cells and retain  
appropriate attenuation. It is useful as a live vaccine for prophylaxis  
of hepatitis A in humans and other primates.

SO Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGFSQAKISLF 20

DB 823 VLPPPRKMGFSQAKISLF 842

RESULT 14

AAP60066  
ID AAP60066 standard; Protein; 2227 AA.

AC AAP60066;

DT 26-JUN-1991 (first entry)

DE Sequence of viral L434 polypeptide encoded by the complete  
nucleotide sequence of the HAV genome.

KM Diagnosis; vaccine; passive immunotherapy.

OS Hepatitis A virus.

PH Key Location/Qualifiers

FT 1..245 /label= P1.1A

FT 246..491 /label= 1B

FT 492..836 /label= 1C

FT 837..980 /label= P2.2A

FT 981..1076 /label= 2B

FT 1077..1422 /label= 2C

FT 1423..1484 /label= P3.3A

FT 1485..1507 /label= 3B

FT 1508..1678 /label= 3C

FT 1679..2227 /label= 3D

PN EPI99480-A.

PD 29-OCT-1986.

PF 03-APR-1986; 86EP-0302465.

PR 03-APR-1985; 85US-0719329.

PA (CHTR-) CHIRON CORP.

PI Dina D, Potter SJ, Vannest GA, Caput D;

DR WPI: 1986-286213/44.

DR N-PSDB; AAN60080.

Hepatitis A virus nucleotide sequence and polypeptide - and use  
in prodn. of vaccines and diagnostic probes

PS Claim 5; Fig 1; 18pp; English.

AAN60080 and oligonucleotide fragments are useful in detection of  
hepatitis A virus; transformed hosts may be used for expression of  
polypeptides and fragments useful in vaccines without risk of  
infection by the virus or in prodn. of particles which are capable  
of inducing immunocompetent B cells for passive immunotherapy. Pref.  
epitope is derived from AAs 445-657 or 792-848 of the HAV  
polypeptide sequence (AAP60066).

SO Sequence 2227 AA;

Query Match 95.0%; Score 96; DB 7; Length 2227;

```

XX  MN9740166-A2.
PN  30-OCT-1997.
XX  18-APR-1997; 97MO-US06506.
XX  19-APR-1996; 96US-0015642.
XX  (USSH ) US SEC DEPT HEALTH.
PA  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX  Emerson SU, Purcell RH, Raychaudhuri G;
XX  WPI: 1997-535850/49.
DR  N-PSDB; AAT93023.
XX  Human attenuated HAV genome containing simian HAV 2C gene - useful
PT  as vaccines against HAV infection
XX  Disclosure: Fig 13A-D; 66pp; English.
XX  This protein sequence is encoded by the human hepatitis A virus
CC  (HAV) HM-175 wild-type genome (see AAT93023). Attenuated strain
CC  HAV/7 is obtained by passage of HM-175 in African Green Monkey
CC  kidney cells. A claimed DNA construct (I) comprises a genome of
CC  HAV, where the genome is a human attenuated HAV genome in which a
CC  region of the 2C gene has been replaced by a corresponding region
CC  from a 2C gene of a simian AGM-27 HAV genome (see AAT93024). The
CC  region of the 2C gene from AGM-27 contained in the construct
CC  preferably encodes amino acids 120-328 of the 2C protein, amino
CC  acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA
CC  transcript of (1); (2) a cell transfected with (1) or the RNA
CC  transcript of (1); (3) a HAV genome as above; (4) antibodies to the
CC  HAV of (3); and (5) a host cell containing the HAV of (3). (1) or
CC  its RNA transcript, can be used as a vaccine for preventing HAV in
CC  a mammal. (1) or the RNA transcript can also be used to stimulate
CC  the production of protective antibodies in the mammal.
XX  Sequence 2227 AA;
SQ

```

Query Match 100.0%; Score 101; DB 18; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY  1 VLPPRRKKMGFLFSQAKISLF 20
    |||||||
DB  823 VLPPRRKKMGFLFSQAKISLF 842

```

RESULT 11  
 ID AAB18607 standard; Protein: 2227 AA.  
 AC AAB18607;  
 DT 15-JAN-2001 (first entry)  
 XX Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 DE HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 KM Hepatitis A virus.  
 XX Hepatitis A virus.  
 OS US6113912-A.  
 PN US6113912-A.  
 XX 05-SEP-2000.  
 PD 07-JUN-1995; 95US-0475886.  
 PF 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93MO-US08610.  
 PS 17-SEP-1993; 93MO-US08610.

```

PR  10-MAR-1995; 95US-0397232.
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA  D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX  WPI: 2000-586464/55.
DR  N-PSDB; AAA75476.
XX  Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT  line useful as vaccine for protecting humans against hepatitis A virus
PT  infection, has modified genome compared to wild type
XX  Disclosure: Fig 6A-K; 72pp; English.
XX  The present sequence is derived from a wild type hepatitis A virus
CC  (HAV) strain HM-174. The sequence is modified to produce HAV which
CC  are adapted to growth in the human fibroblast-like cell line MRC-5.
CC  The HAV is able to propagate in MRC-5 cells and retain appropriate
CC  attenuation. It is useful as a live vaccine for prophylaxis of
CC  hepatitis A in humans and other primates.
XX  Sequence 2227 AA;
SQ

```

Query Match 100.0%; Score 101; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY  1 VLPPRRKKMGFLFSQAKISLF 20
    |||||||
DB  823 VLPPRRKKMGFLFSQAKISLF 842

```

RESULT 12  
 ID AAB18608 standard; Protein: 2227 AA.  
 AC AAB18608;  
 DT 15-JAN-2001 (first entry)  
 XX Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
 DE HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KM P-35 virus.  
 XX Hepatitis A virus.  
 OS US6113912-A.  
 PN US6113912-A.  
 XX 05-SEP-2000.  
 PD 07-JUN-1995; 95US-0475886.  
 PF 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93MO-US08610.  
 PR 10-MAR-1995; 95US-0397232.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX WPI: 2000-586464/55.  
 DR N-PSDB; AAA75477.  
 XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type  
 XX Disclosure: Columns 67-78; 72pp; English.  
 CC The present sequence is derived from passage 35 of a wild type





PN EPI38704-A.  
XX  
XX 24-APR-1985.  
XX  
XX 09-OCT-1984; 84EP-0402025.  
XX  
XX 02-MAR-1984; 84US-0585942.  
XX 14-OCT-1983; 83US-0541836.  
XX  
XX (MERI ) MERCK & CO INC.  
XX  
XX Hughes JV, Scolnick EM, Tomassini JE;  
XX  
XX WPI; 1985-100818/17.  
XX N-PSDB; AAN50274.  
XX  
XX New hepatitis A virus surface protein - useful for binding to  
XX neutralising antibodies to the virus  
XX  
XX Disclosure; Page 17-23; 49pp; English.  
XX  
XX VPI is isolated by solubilisation of the intact virus in an aq.  
XX antionic surfactant and a reducing agent. The viral proteins are sepd.  
XX and the protein of molecular wt. 33000 daltons is sepd.  
XX  
XX Sequence 993 AA;  
SQ

Query Match 100.0%; Score 101; DB 6; Length 993;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20  
Db 959 VLPPRRKMKGLFSQAKISLF 978

RESULT 7  
AAW95559  
ID AAW95559 standard; Protein; 1077 AA.  
XX  
XX AAW95559;  
AC  
XX  
XX 28-APR-1999 (first entry)  
DT  
XX  
XX A partial hepatitis A virus (HAV) protein.  
DE  
XX  
XX Hepatitis A virus protein; HAV; P2 region;  
KM cell-culture-adapted HAV strain; infection; accelerated growth.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX US5849562-A.  
PN  
XX  
XX 15-DEC-1998.  
PD  
XX  
XX 06-JUN-1995; 95US-0468926.  
PF  
XX  
XX 06-NOV-1991; 91US-0788262.  
PR 30-SEP-1983; 83US-0537911.  
PR 27-SEP-1984; 84US-0654942.  
PR 06-OCT-1988; 88US-0256135.  
PR 06-JUN-1995; 95US-0468926.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Emerson ST, Purcell RH;  
XX  
XX WPI; 1999-094412/08.  
XX N-PSDB; AAX01006.  
XX  
XX Chimeric hepatitis A virus strains - with P2 region from  
XX cell-culture-adapted strain in wild-type genome

XX  
XX Disclosure; Fig 7A-L; 36pp; English.  
PS  
XX  
XX The present sequence represents a partial hepatitis A virus (HAV)  
CC protein. The specification describes a DNA construct consisting  
CC of a wild-type HAV genome in which the P2 region is replaced by the  
CC P2 region from a cell-culture-adapted HAV strain. The construct is  
CC used to demonstrate that mutations in the P2 region of a  
CC cell-culture-adapted HAV strain are sufficient for establishment of  
CC infection and accelerated growth in cell culture.  
XX  
XX Sequence 1077 AA;  
SQ

Query Match 100.0%; Score 101; DB 20; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPRRKMKGLFSQAKISLF 20  
Db 1046 VLPPRRKMKGLFSQAKISLF 1065

RESULT 8  
AAR32426  
ID AAR32426 standard; Protein; 1091 AA.  
XX  
XX AAR32426;  
AC  
XX  
XX 17-DEC-2001 (updated)  
DT 10-JUN-1993 (first entry)  
DT  
XX  
XX Translated from 5' region of Hepatitis A virus genomic clone.  
DE  
XX  
XX HAV HM-175; chronic liver disease; picornavirus.  
KM  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX  
XX  
XX Key Location/Qualifiers  
FH 238..1091  
FT /label= ORF  
FT /note= "second putative initiation codon at  
FT position 240"  
FT  
FT Region 1..711  
FT /note= "X's correspond to nonsense codons,  
FT i.e. this region is not an ORF"  
XX  
XX  
XX USN788262-N.  
PN  
XX  
XX 15-DEC-1992.  
PD  
XX  
XX 30-SEP-1983; 83US-0536911.  
PF  
XX  
XX 27-SEP-1984; 84US-0654942.  
PR 06-OCT-1988; 88US-0256135.  
PR 30-SEP-1983; 83US-0536911.  
PR 06-NOV-1991; 91US-0788262.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
XX Baltimore D, Feinstone SM;  
XX PI Purcell RH, Racanelli VR, Ticehurst JR;  
XX  
XX WPI; 1993-067429/08.  
XX N-PSDB; AAO36934.  
XX  
XX  
XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.  
XX of antigen and antibodies  
XX  
XX Disclosure; Fig 7; 65pp; English.  
PS  
XX  
XX HAV virion RNA was extracted from the livers of marmosets which had  
XX been inoculated with HAV (the HAV had previously been passaged twice

Query Match 100.0%; Score 101; DB 6; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGKLFQSAKISLF 20  
 |||  
 DB 332 VLPPPRKMGKLFQSAKISLF 351

#### RESULT 4

AAP50287  
 ID AAP50287 standard; Protein; 854 AA.

AC AAP50287;

DT 30-NOV-1991 (first entry)

DE Sequence encoded by hepatitis A virus (HAV) cDNA from near the  
 DE genome 5' terminus to the end of the area corresponding to the  
 DE capsid protein region of poliovirus RNA.

KM Hepatitis A virus assay; antigen; antibody.

OS Hepatitis A virus.

PN W08501517-A.

PD 11-APR-1985.

PF 27-SEP-1984; 84WO-US01552.

PR 30-SEP-1983; 83OS-0537911.

PA (MASI ) MASSACHUSETTS INST TECH.

PI Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;

PI Racanelli VR;

DR WPI; 1985-09846/16.

DR N-PSDB; AAN50330.

PT New hepatitis A virus cDNA - useful in assays for the virus and  
 for prodn. of the viral antigen and antibodies to it

PS Example; Fig 7: 60pp; English.

CC The inventors claim HAV cDNA and a method for producing it, whereby  
 CC large ants. can be obt'd. economically. The cDNA is useful in the  
 CC assay for detection of HAV quickly and easily and with high  
 CC sensitivity and specificity. The HAV cDNA is also used in the prodn.  
 CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.

SO Sequence 854 AA;

Query Match 100.0%; Score 101; DB 6; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGKLFQSAKISLF 20  
 |||  
 DB 823 VLPPPRKMGKLFQSAKISLF 842

#### RESULT 5

AAP50116  
 ID AAP50116 standard; Protein; 993 AA.

AC AAP50116;

DT 30-SEP-1991 (first entry)

DE Sequence of Hepatitis A virus (HAV) immunogenic peptides  
 DE VP-1, VP-2, VP-3 and VP-4.

KM Antigenic protein; immunogen; vaccine.

OS Hepatitis A virus (strain CR326).

PN BP154587-A.

PD 11-SEP-1985.

PF 27-FEB-1985; 85EP-0400369.

PR 02-MAR-1984; 84US-0585818.

PA (MERI ) MERCK & CO INC.

PI Linemeyer DL, Menke JG, Reuben RG, Mitra SM;

DR WPI; 1985-224964/37.

DR N-PSDB; AAN50139.

PT New nucleotide sequences coding for hepatitis A virus antigens -  
 PT useful for eliciting normal immune response and in vaccines for  
 PT protecting against the virus

PS Example; Page 11-17; 32pp; English.

CC Within the sequence in AAN50139 is encoded the information necessary  
 CC to make the antigenic proteins of HAV. The sequences encoding for  
 CC the structural proteins begin at base 403. The key sub-unit  
 CC sequences within VP-1, designated Sequences I, II, III, IV, and V,  
 CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other  
 CC nucleotide sequences which are valuable as encoding antigenic  
 CC proteins are the sequences from base 1749 to base 2722; from base  
 CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from  
 CC base 1749 to base 2722 is esp. valuable as a vector for producing  
 CC antigen protein. Sequences II-V are claimed. X in AAP50116 denotes the  
 CC translation of a stop codon.

SO Sequence 993 AA;

Query Match 100.0%; Score 101; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMGKLFQSAKISLF 20  
 |||  
 DB 959 VLPPPRKMGKLFQSAKISLF 978

#### RESULT 6

AAP50231  
 ID AAP50231 standard; Protein; 993 AA.

AC AAP50231;

DT 28-NOV-1991 (first entry)

DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
 DE including surface protein (VP-1).

KM Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KM diagnostic assay.

OS Hepatitis A virus.

Key Location/Qualifiers

FT Protein

FT 628..993  
 FT /note="Claimed: X denotes translated stop codons  
 and unspecified triplets"

XX	18-APR-1997;	97WO-US06891.
PF	19-APR-1996;	96US-0015644.
PR	(USSH )	US DEPT HEALTH & HUMAN SERVICES.
XX	Fields HA,	Khudyakov YE;
PA	WPI; 1997-535831/49.	
XX	Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an	
DR	immune response to HAV in a mammal or to detect the presence of	
PT	antibodies against HAV in a mammal	
XX	Claim 16: Page 112; 140pp; English.	
PS		
XX	Peptides AAW42922-30 are immunogenic peptides corresponding to	
CC	immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are	
CC	substantially similar to a portion of the amino acid sequence of the P2A	
CC	protein of HAV corresponding to amino acids 792-980. The present peptide	
CC	is derived from amino acids 823-842, and has a reactivity of 31.3% with	
CC	acute sera. Compositions containing the peptides can be used to induce an	
CC	immune response to HAV in a mammal. The peptides can also be used to	
CC	detect the presence of antibodies against HAV in mammalian serum. The	
CC	peptides can also be used to make an antibody against HAV by	
CC	administering the peptide to a mammal.	
XX		
XX	Sequence 20 AA;	
XX		

Query Match	100.0%;	Score 101;	DB 18;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 2, 5e-09;		
Matches	20;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0
OY	1	VLPPEPRKMGLEFSQAKISLF	20	
Db	1	VLPPEPRKMGLEFSQAKISLF	20	
RESULT 2				
AAB69442	ID	AAB69442 standard; Peptide: 21 AA.		
XX	AC	AAB69442;		
XX	20-APR-2001	(first entry)		
XX	Synthetic	HAV P2A peptide, SEQ ID NO: 42.		
XX	Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;			
KW	antigen; major structural capsid polypeptide; HAV antibody detection.			
XX	Hepatitis A virus.			
OS	Synthetic.			
XX	MO200105824-A2.			
XX	25-JAN-2001.			
XX	14-JUL-2000;	2000MO-US19267.		
XX	15-JUL-1999;	99US-0144412.		
XX	(USSH )	US DEPT HEALTH & HUMAN SERVICES.		
XX	Fields HA,	Khudyakov YE;		
XX	PI	WPI; 2001-112681/12.		
XX	Synthetic peptides used as antigen sources for enzyme immunoassays			
XX	detecting anti-hepatitis A virus and as vaccines -			

PS Claim 13, Page 95, 130pp, English.

```

Query Match: 100.0%; Score 101; DB 22; Length 21;
Best local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps
QY      1 VLPPPRKMGLEFSQAQISLP 20
        |||
Db       1 VLPPPTKMGLTSGAKISLF 20

RESULT   3
AAP50230
AAP50230 standard; Protein: 366 AA.
XX
AC      AAP50230;
XX
DT      28-NOV-1991 (first entry)
XX
DE      Sequence of hepatitis A virus (HAV) surface protein (VP-1).
XX
KM      Hepatitis A virus vaccine; Immunisation; monoclonal antibody;
XX      diagnostic assay.
XX
OS      Hepatitis A virus.
XX
EP138704-A.
XX
PD      24-APR-1985.
XX
PF      09-OCT-1984; 84EP-0402025.
XX
PR      02-MAR-1984; 84US-0565942.
XX      14-OCT-1983; 83US-0541836.
PA      (MERI ) MERCK & CO INC.
XX
PI      Hughes JV, Scolnick EM, Tomassini JE;
DR      WPI; 1985-100818/17.
DR      N-PSTDB; AAN50274.
XX
PT      New hepatitis A virus surface protein - useful for binding to
PT      neutralising antibodies to the virus
XX
PS      Claim 21; Page 46-48; 49pp; English.
XX
CC      VP1 is isolated by solubilisation of the intact virus in an aq.
CC      antionic surfactant and a reducing agent. The viral proteins are sepd
CC      and the protein of molecular wt. 33000 daltons is sepd.
XX
SQ      Sequence 366 AA;

```





DR EMBL: AB038298; BAB1836.1; -  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10807 MW; F194CE91BB8C4AFD CRC64;

Query Match  
Best Local Similarity 100.0%; Score 103; DB 12; Length 94;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 61 ORLKVAQEELSNEVLPPPRK 80

## RESULT 23

ID Q9ENT8 PRELIMINARY; PRT; 94 AA.  
AC Q9ENT8:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ACC 1:  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038299; BAB1837.1; -  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;

Query Match  
Best Local Similarity 100.0%; Score 103; DB 12; Length 94;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 61 ORLKVAQEELSNEVLPPPRK 80

## RESULT 24

ID Q9ENT7 PRELIMINARY; PRT; 94 AA.  
AC Q9ENT7:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ACC 2:  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038300; BAB1838.1; -  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;

SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;

Query Match  
Best Local Similarity 100.0%; Score 103; DB 12; Length 94;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 61 ORLKVAQEELSNEVLPPPRK 80

## RESULT 25

ID Q9ENT6 PRELIMINARY; PRT; 94 AA.  
AC Q9ENT6:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ACC 20:  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038301; BAB1839.1; -  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;

Query Match  
Best Local Similarity 100.0%; Score 103; DB 12; Length 94;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 61 ORLKVAQEELSNEVLPPPRK 80

Search completed: June 16, 2002, 00:08:52  
Job time: 790 sec

RC STRAIN-NAGASAKI 27;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038294; BAB11832.1; -  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10892 MW; 9AF8BE91BB8C53CC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPPRK 20  
Db 61 ORLKYAOEELSNEVLPPPRK 80

RESULT 19  
O9ENDU2 PRELIMINARY; PRT; 94 AA.  
AC O9ENDU2:  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 28;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL; AB038295; BAB11833.1; -  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPPRK 20  
Db 61 ORLKYAOEELSNEVLPPPRK 80

RESULT 20  
O9ENDU1 PRELIMINARY; PRT; 94 AA.  
AC O9ENDU1:  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 29;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;

RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038296; BAB11834.1; -  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPPRK 20  
Db 61 ORLKYAOEELSNEVLPPPRK 80

RESULT 21  
O9ENDU0 PRELIMINARY; PRT; 94 AA.  
AC O9ENDU0:  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 30;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL; AB038297; BAB11835.1; -  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10844 MW; 9AF8A263BB895FAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPPRK 20  
Db 61 ORLKYAOEELSNEVLPPPRK 80

RESULT 22  
O9END9 PRELIMINARY; PRT; 94 AA.  
AC O9END9:  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 32;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."

OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RN NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 23;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038290; BAB1828.1; -.  
FT NON\_TER 1 1  
FT SEQUENCE 94 AA: 10876 MW: 9AF8BE91BB8C4BAC CRC64;  
SQ

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNEVLPPPK 20  
DB 61 ORLKTAQEELSNEVLPPPK 80

RESULT 15  
OGENU6 PRELIMINARY; PRT; 94 AA.  
ID OGENU6:  
AC OGENU6: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RN NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 24;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038291; BAB1829.1; -.  
FT NON\_TER 1 1  
FT SEQUENCE 94 AA: 10876 MW: 9AF8BE91BB8C4BAC CRC64;  
SQ

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNEVLPPPK 20  
DB 61 ORLKTAQEELSNEVLPPPK 80

RESULT 16  
OGENU5 PRELIMINARY; PRT; 94 AA.  
ID OGENU5:  
AC OGENU5: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RN NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.

OX NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 25;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038292; BAB1830.1; -.  
FT NON\_TER 1 1  
FT SEQUENCE 94 AA: 10876 MW: 9AF8BE91BB8C4BAC CRC64;  
SQ

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNEVLPPPK 20  
DB 61 ORLKTAQEELSNEVLPPPK 80

RESULT 17  
OGENU4 PRELIMINARY; PRT; 94 AA.  
ID OGENU4:  
AC OGENU4: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RN NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 26;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038293; BAB1831.1; -.  
FT NON\_TER 1 1  
FT SEQUENCE 94 AA: 10876 MW: 9AF8BE91BB8C4BAC CRC64;  
SQ

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNEVLPPPK 20  
DB 61 ORLKTAQEELSNEVLPPPK 80

RESULT 18  
OGENU3 PRELIMINARY; PRT; 94 AA.  
ID OGENU3:  
AC OGENU3: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RN NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.



ID 09ENV1 PRELIMINARY; PRT; 94 AA.  
AC 09ENV1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DE 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 18;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB038286; BAB11824.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNVLP PPRK 20  
Db 61 ORLKTAQEELSNVLP PPRK 80

RESULT 11  
09ENV0 PRELIMINARY; PRT; 94 AA.  
AC 09ENV0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DE 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 20;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB038287; BAB11825.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNVLP PPRK 20  
Db 61 ORLKTAQEELSNVLP PPRK 80

RESULT 12  
09ENV9 PRELIMINARY; PRT; 94 AA.  
AC 09ENV9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 21;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB038288; BAB11826.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNVLP PPRK 20  
Db 61 ORLKTAQEELSNVLP PPRK 80

RESULT 13  
09ENV8 PRELIMINARY; PRT; 94 AA.  
AC 09ENV8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DE 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 22;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB038289; BAB11827.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKTAQEELSNVLP PPRK 20  
Db 61 ORLKTAQEELSNVLP PPRK 80

RESULT 14  
09ENV7 PRELIMINARY; PRT; 94 AA.  
AC 09ENV7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DE 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).

OY 1 ORLKVAOELSNEVLPPRK 20  
Db 61 ORLKVAOELSNEVLPPRK 80

## RESULT 6

ID O9ENV6 PRELIMINARY; PRT; 94 AA.  
AC O9ENV6;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 07;  
RA Ida S.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV) infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038281; BAB11819.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAOELSNEVLPPRK 20  
Db 61 ORLKVAOELSNEVLPPRK 80

## RESULT 7

ID O9ENV5 PRELIMINARY; PRT; 94 AA.  
AC O9ENV5;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 08;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV) infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038282; BAB11820.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAOELSNEVLPPRK 20  
Db 61 ORLKVAOELSNEVLPPRK 80

RESULT 8  
ID O9ENV4 PRELIMINARY; PRT; 94 AA.  
AC O9ENV4;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 10;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV) infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038283; BAB11821.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAOELSNEVLPPRK 20  
Db 61 ORLKVAOELSNEVLPPRK 80

## RESULT 9

ID O9ENV2 PRELIMINARY; PRT; 94 AA.  
AC O9ENV2;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 15;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV) infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038285; BAB11823.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKVAOELSNEVLPPRK 20  
Db 61 ORLKVAOELSNEVLPPRK 80

## RESULT 10

O9ENV1

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPrK 20  
|||||  
DB 61 ORLKYAOEELSNEVLPPrK 80

## RESULT 2

OQENV0 PRELIMINARY: PRT: 94 AA.  
AC OQENV0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 02;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT \*Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038277; BAB11815.1; -.  
FT NON\_TER 1 94  
FT SEQUENCE 94 AA; 10903 MW; 9AF8BE853D57329C CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPrK 20  
|||||  
DB 61 ORLKYAOEELSNEVLPPrK 80

## RESULT 3

OQENV9 PRELIMINARY: PRT: 94 AA.  
AC OQENV9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 03;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT \*Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038278; BAB11816.1; -.  
FT NON\_TER 1 94  
FT SEQUENCE 94 AA; 10862 MW; 9AF9EFD4A8C4BAC CRC64;

Query Match

100.0%; Score 103; DB 12; Length 94;

Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPrK 20  
|||||  
DB 61 ORLKYAOEELSNEVLPPrK 80

## RESULT 4

OQENV8 PRELIMINARY: PRT: 94 AA.  
AC OQENV8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 04;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT \*Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038279; BAB11817.1; -.  
FT NON\_TER 1 94  
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPrK 20  
|||||  
DB 61 ORLKYAOEELSNEVLPPrK 80

## RESULT 5

OQENV7 PRELIMINARY: PRT: 94 AA.  
AC OQENV7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 06;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT \*Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB038280; BAB11818.1; -.  
FT NON\_TER 1 94  
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:51 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-41

Perfect score: 103

Sequence: 1 ORLKYAQEELSNVLPFPRK 20

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_PBC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VIRIBACTERIA:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	94	09ENW1	09enw1 hepatitis a
2	103	100.0	94	09ENW2	09enw2 hepatitis a
3	103	100.0	94	09ENW3	09enw3 hepatitis a
4	103	100.0	94	09ENW4	09enw4 hepatitis a
5	103	100.0	94	09ENW5	09enw5 hepatitis a
6	103	100.0	94	09ENW6	09enw6 hepatitis a
7	103	100.0	94	09ENW7	09enw7 hepatitis a
8	103	100.0	94	09ENW8	09enw8 hepatitis a
9	103	100.0	94	09ENW9	09enw9 hepatitis a
10	103	100.0	94	09ENW10	09enw10 hepatitis a
11	103	100.0	94	09ENW11	09enw11 hepatitis a
12	103	100.0	94	09ENW12	09enw12 hepatitis a
13	103	100.0	94	09ENW13	09enw13 hepatitis a
14	103	100.0	94	09ENW14	09enw14 hepatitis a
15	103	100.0	94	09ENW15	09enw15 hepatitis a
16	103	100.0	94	09ENW16	09enw16 hepatitis a

## ALIGNMENTS

17	103	100.0	94	12	09ENW4	09enw4 hepatitis a
18	103	100.0	94	12	09ENW3	09enw3 hepatitis a
19	103	100.0	94	12	09ENW2	09enw2 hepatitis a
20	103	100.0	94	12	09ENW1	09enw1 hepatitis a
21	103	100.0	94	12	09ENW0	09enw0 hepatitis a
22	103	100.0	94	12	09ENW9	09enw9 hepatitis a
23	103	100.0	94	12	09ENW8	09enw8 hepatitis a
24	103	100.0	94	12	09ENW7	09enw7 hepatitis a
25	103	100.0	94	12	09ENW6	09enw6 hepatitis a
26	103	100.0	94	12	09ENW5	09enw5 hepatitis a
27	103	100.0	94	12	09ENW4	09enw4 hepatitis a
28	103	100.0	94	12	09ENW3	09enw3 hepatitis a
29	103	100.0	94	12	09ENW2	09enw2 hepatitis a
30	103	100.0	94	12	09ENW1	09enw1 hepatitis a
31	103	100.0	94	12	09ENW0	09enw0 hepatitis a
32	103	100.0	94	12	09ENW9	09enw9 hepatitis a
33	103	100.0	94	12	09ENW8	09enw8 hepatitis a
34	103	100.0	94	12	09ENW7	09enw7 hepatitis a
35	103	100.0	94	12	09ENW6	09enw6 hepatitis a
36	103	100.0	94	12	09ENW5	09enw5 hepatitis a
37	103	100.0	94	12	09ENW4	09enw4 hepatitis a
38	103	100.0	94	12	09ENW3	09enw3 hepatitis a
39	103	100.0	94	12	09ENW2	09enw2 hepatitis a
40	103	100.0	94	12	09ENW1	09enw1 hepatitis a
41	103	100.0	94	12	09ENW0	09enw0 hepatitis a
42	103	100.0	94	12	09ENW9	09enw9 hepatitis a
43	103	100.0	94	12	09ENW8	09enw8 hepatitis a
44	103	100.0	94	12	09ENW7	09enw7 hepatitis a
45	103	100.0	94	12	09ENW6	09enw6 hepatitis a
46	103	100.0	94	12	09ENW5	09enw5 hepatitis a
47	103	100.0	94	12	09ENW4	09enw4 hepatitis a
48	103	100.0	94	12	09ENW3	09enw3 hepatitis a
49	103	100.0	94	12	09ENW2	09enw2 hepatitis a
50	103	100.0	94	12	09ENW1	09enw1 hepatitis a
51	103	100.0	94	12	09ENW0	09enw0 hepatitis a
52	103	100.0	94	12	09ENW9	09enw9 hepatitis a
53	103	100.0	94	12	09ENW8	09enw8 hepatitis a
54	103	100.0	94	12	09ENW7	09enw7 hepatitis a
55	103	100.0	94	12	09ENW6	09enw6 hepatitis a
56	103	100.0	94	12	09ENW5	09enw5 hepatitis a
57	103	100.0	94	12	09ENW4	09enw4 hepatitis a
58	103	100.0	94	12	09ENW3	09enw3 hepatitis a
59	103	100.0	94	12	09ENW2	09enw2 hepatitis a
60	103	100.0	94	12	09ENW1	09enw1 hepatitis a

RESULT 1  
09ENW1 PRELIMINARY; PRT; 94 AA.  
AC 09ENW1;  
DT 01-MAR-2001 (TREMBLERel. 16, Created)  
DT 01-MAR-2001 (TREMBLERel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLERel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 01;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038276; BAB11814.1;  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

**THIS PAGE BLANK (USPTO)**



RL Hum. Mol. Genet. 4:837-842(1995).  
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE  
CC PLASMA MEMBRANE.  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 11 DISTINCT ISOFORMS EXIST.  
CC -1- TISSUE SPECIFICITY: DIFFERENTIALLY EXPRESSED DURING SKELETAL  
CC MUSCLE, HEART, AND BRAIN DEVELOPMENT. ALSO EXPRESSED IN  
CC RETINA.  
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY  
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,  
CC ABP-120, ABP-180, OR BETA-FODRIN).  
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M68859; AAB02797.1; -  
DR EMBL: M18025; AAA37530.1; -  
DR EMBL: U56724; AAB01216.1; -  
DR EMBL: U15218; AAA87068.1; -  
DR PIR: B27162; B27162.  
DR HSP: P46939; 1BHD.  
DR MGD: MGI:94909; Dmd.  
DR InterPro: IPR001589; Actinin\_act\_bind.  
DR InterPro: IPR001715; Calponin\_hom.  
DR InterPro: IPR002017; Spectrin.  
DR InterPro: IPR001202; WW.  
DR InterPro: IPR002349; WW\_domain.  
DR InterPro: IPR000433; Znf\_zz.  
DR Pfam: PF00307; CH; 2.  
DR Pfam: PF00435; Spectrin; 21.  
DR Pfam: PF00397; WW; 1.  
DR Pfam: PF00569; Z2; 1.  
DR PRINTS: PR00403; WMDOMAIN.  
DR SMART: SM00033; CH; 2.  
DR SMART: SM00150; SPEC; 20.  
DR SMART: SM00456; WW; 1.  
DR SMART: SM00291; Znf\_zz; 1.  
DR PROSITE: PS00019; ACTININ\_1; 1.  
DR PROSITE: PS00020; ACTININ\_2; 1.  
DR PROSITE: PS50021; CH; 2.  
DR PROSITE: PS01159; WW\_DOMAIN\_1; 1.  
DR PROSITE: PS50020; WW\_DOMAIN\_2; 1.  
DR PROSITE: PS01357; ZF\_ZZ\_1; 2.  
DR PROSITE: PS50135; ZF\_ZZ\_2; 1.  
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;  
KW Repeat; Alternative splicing; Zinc-finger.  
FT DOMAIN 1 240  
FT DOMAIN 15 119 CH 1.  
FT DOMAIN 134 237 CH 2.  
FT REPEAT 341 449 SPECTRIN 1.  
FT REPEAT 450 558 SPECTRIN 2.  
FT REPEAT 561 669 SPECTRIN 3.  
FT REPEAT 721 830 SPECTRIN 4.  
FT REPEAT 832 936 SPECTRIN 5.  
FT REPEAT 945 1047 SPECTRIN 6.  
FT REPEAT 1050 1156 SPECTRIN 7.  
FT REPEAT 1159 1265 SPECTRIN 8.  
FT REPEAT 1268 1369 SPECTRIN 9.  
FT REPEAT 1470 1570 SPECTRIN 10.  
FT REPEAT 1573 1678 SPECTRIN 11.  
FT REPEAT 1681 1782 SPECTRIN 12.  
FT REPEAT 1879 1981 SPECTRIN 13.  
FT REPEAT 2013 2103 SPECTRIN 14.  
FT REPEAT 2106 2210 SPECTRIN 15.  
FT REPEAT 2213 2318 SPECTRIN 16.

FT	REPEAT	2468	2570	SPECTRIN 17.
FT	REPEAT	2573	2679	SPECTRIN 18.
FT	REPEAT	2682	2795	SPECTRIN 19.
FT	REPEAT	2798	2900	SPECTRIN 20.
FT	REPEAT	2902	2934	SPECTRIN 21.
FT	REPEAT	2927	3033	SPECTRIN 22.
FT	DOMAIN	3048	3081	WW.
FT	ZN_FING	3300	3347	Z2-TYPE.
FT	CONFLICT	463	463	D -> H (IN REF. 3).
FT	CONFLICT	677	677	S -> F (IN REF. 3).
FT	CONFLICT	2337	2337	V -> L (IN REF. 1; AAB02797).
SO	SEQUENCE	3678 AA:	425810 MM:	1D2E74CF7DB035EE CRC64;

Query Match 42.2%; Score 43.5; DB 1; Length 3678;  
Best Local Similarity 55.6%; Pred. No. 2.7e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
QY 3 LKTAQBELSNEVLPPPK 20  
:|:||||| |:  
Db 698 VKHAQELPP---PPPK 712

Search completed: June 16, 2002, 00:10:07  
Job time: 655 sec





CC -1- SIMILARITY: TO ANYLPHORINS AND TO ARTHROPOD HEMOCYANINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: U01338; AAA74579.1; -  
 CC DR HSSP: P04253; 10XY.  
 CC DR InterPro: IPR000896; Hemocyanin.  
 CC DR Pfam: PF00372; hemocyanin; 2.  
 CC DR PRINTS: PR00187; HAEMOCYANIN.  
 CC DR PROSITE: PS00209; HEMOCYANIN.1; FALSE\_NEG.  
 CC DR PROSITE: PS00210; HEMOCYANIN.2; FALSE\_NEG.  
 CC KW Signal; storage protein; glycoprotein.  
 CC FT SIGNAL 1 17  
 CC FT CHAIN 1 17  
 CC FT CARBOHYD 18 733  
 CC FT CARBOHYD 199 199  
 CC FT CARBOHYD 234 234  
 CC FT CARBOHYD 431 431  
 CC SO SEQUENCE 733 AA; 87813 MW; 083DF739DD65729 CRC64;

Query Match 43.7%; Score 45; DB 1; Length 733;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 5 YAOEELSNEVLPP 18  
 DB 148 FORDLSYLLPPP 161

RESULT 22  
 ID PMG3\_YEAST STANDARD: PRT: 303 AA.  
 AC Q12326;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Phosphoglycerate mutase 3 (EC 5.4.2.1) (Phosphoglyceromutase 3) (PGAM  
 3) (MPCM 3) (BPG-dependent PGAM 3).  
 GN GPM3 OR YOL056W OR Q1236.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / FY73;  
 RX MEDLINE=9820582; PubMed=8789261;  
 RA Mannhaupt G., Vetter I., Schwarzlose C., Mitzel S., Feldmann H.;  
 RT "Analysis of a 26 kb region on the left arm of yeast chromosome XV";  
 RT yeast 12:67-76(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ansonger W., Benes V., Rechmann S., Schwager C., Teodoru C., Voss H.,  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION.  
 RA MEDLINE=98205882; PubMed=9544241;  
 RA Hehlisch J.J., Mueller S., Schueter E., Jacoby J., Rodicio R.;  
 RT "Investigation of two yeast genes encoding putative isoenzymes of  
 RT phosphoglycerate mutase";  
 RT yeast 14:203-213(1998).  
 CC -1- FUNCTION: COULD BE NON-FUNCTIONAL.  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate  
 CC = 3-phospho-D-glycerate + 2,3-diphosphoglycerate.  
 CC -1- PATHWAY: GLYCOLYSIS.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE MUTASE FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X91067; CAA62530.1; -  
 CC DR EMBL: 274798; CAA99064.1; -  
 CC DR HSSP: P00950; 5PGM.  
 CC DR SGD: S0003417; GPM3.  
 CC DR InterPro: IPR001345; PG\_mutase.  
 CC DR Pfam: PF00300; PGAM; 3.  
 CC DR PROSITE: PS00175; PG\_MUTASE; 1.  
 CC KW Isomerase; Glycolysis.  
 CC FT ACT\_SITE 14 14  
 CC FT ACT\_SITE 14 14  
 CC FT ACT\_SITE 70 70  
 CC FT ACT\_SITE 235 235  
 CC FT ACT\_SITE 235 235  
 CC SO SEQUENCE 303 AA; 34863 MW; 29C3F3D28560914 CRC64;

Query Match 42.7%; Score 44; DB 1; Length 303;  
 Best Local Similarity 56.2%; Pred. No. 14;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 QRLKYOEELSNEVL 16  
 DB 183 RHLKYGPEERANRLP 198

RESULT 23  
 ID PMG1\_ANTSP STANDARD: PRT: 510 AA.  
 AC Q06464;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Putative 2,3-bisphosphoglycerate-independent phosphoglycerate mutase  
 DE (EC 5.4.2.1) (phosphoglyceromutase) (BPG-independent PGAM).  
 GN PGAM.  
 OS Antithamnion sp.  
 OC Chloroplast.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales;  
 CC Anthamion.  
 OX NCBI\_TaxID=2767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93173098; PubMed=8437571;  
 RA Valentin K.U.;  
 RT "Sech is plastid-encoded in a red alga: implications for the  
 RT evolution of plastid genomes and the thylakoid protein import  
 RT apparatus";  
 RT Mol. Gen. Genet. 236:245-250(1993).  
 CC -1- FUNCTION: KEY ENZYME OF PRIMARY CARBON METABOLISM (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-PHOSPHOGLYCERATE -> 3-PHOSPHOGLYCERATE.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast.  
 CC -1- SIMILARITY: HIGH TO OTHER BPG-INDEPENDENT PGAM; SOME, TO THE  
 CC ALKALINE PHOSPHATASE FAMILY OF ENZYMES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X64705; CAA45959.1; -  
 CC PIR: S42705; S42705.

RX MEDLINE-91065520: PubMed-2123466:  
RA Hu Y.-F., Luescher B., Admon A., Mermod N., Tian R.:  
RT "Transcription factor AP-4 contains multiple dimerization domains  
that regulate dimer specificity".  
RL Genes Dev. 4:1741-1752(1990).  
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT ACTIVATES BOTH VIRAL AND  
CELLULAR GENES BY BINDING TO THE SYMMETRICAL DNA SEQUENCE  
5'-CACATC-3'.  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC BHLH PROTEIN, HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: S73885; AAB32235.1; -  
DR EMBL: AC004653; AAC17116.1; -  
DR EMBL: X57435; CAA40683.1; -  
DR PIR: A36394; A36394.  
DR HSSP: P25912; IHLO.  
DR TRASNFCAC: T00036; -  
DR MIM: 600743; -  
DR InterPro: IPR003015; HLH\_MYC.  
DR InterPro: IPR001092; HLH\_dim.  
DR Pfam: PF00010; HLH; 1.  
DR SMART: SM00353; HLH; 1.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.  
FT DNA\_BIND 48 60 BASIC DOMAIN.  
FT DOMAIN 61 100 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 99 120 LEUCINE-ZIPPER 1.  
FT DOMAIN 131 179 LEUCINE-ZIPPER 2.  
FT DOMAIN 193 222 GLN-RICH.  
FT DOMAIN 225 244 PRO-RICH.  
FT CONFLICT 218 218 O -> H (IN REF. 2).  
SQ SEQUENCE 338 AA; 38725 MW; 540C008558596883 CRC64;

Query Match 44.7%; Score 46; DB 1; Length 338;  
Best Local Similarity 38.9%; Pred. No. 7.4;  
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELNEVLP 18  
DB 210 EKLREDOOLRTOLP 227  
::: ||:| ::|||

RESULT 20  
ID 11SB\_CUCMA STANDARD: PRT; 480 AA.  
AC P13744;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE 11S globulin beta subunit precursor.  
OS Cucurbita maxima (Pumpkin) (Winter squash).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustosids I; Cucurbitales; Cucurbitaceae; Cucurbita.  
OX NCBI\_TaxID=3661;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. KUROKAWA AMAKURI NANKIN;  
RX MEDLINE-88166744; PubMed-2450746;  
RA Hayashi M., Mori H., Nishimura M., Akazawa T., Hara-Nishimura I.;  
RT "Nucleotide sequence of cloned cDNA coding for pumpkin 11-S globulin

RT beta subunit.";  
RL Eur. J. Biochem. 172:627-632(1988).  
RN [2]  
RP SEQUENCE OF 22-30 AND 297-302.  
RA Ohmura M., Hara I., Mastubara H.;  
RT "Pumpkin (Cucurbita sp.) seed globulin IV. Terminal sequences of the  
RT acidic and basic peptide chains and identification of a pyroglutamy  
RT peptide chain".  
RL Plant Cell Physiol. 21:157-167(1980).  
CC -1- FUNCTION: THIS IS A SEED STORAGE PROTEIN.  
CC -1- SUBUNIT: HEXAMER. EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A  
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A  
CC DISULFIDE BOND.  
CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEIN (GLOBULINS)  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M36407; AAA33110.1; -  
DR PIR: S00366; FWPUB.  
DR InterPro: IPR000459; Seedstore\_11s.  
DR Pfam: PF00190; Seedstore\_11s; 1.  
DR PRINTS: PR00439; 11SGLOBULIN.  
DR PROSITE: PS00305; 11S\_SEED\_STORAGE; 1.  
KW Seed storage protein; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 480 11S GLOBULIN BETA SUBUNIT.  
FT CHAIN 22 296 GAMMA CHAIN (ACIDIC).  
FT CHAIN 297 480 DELTA CHAIN (BASIC).  
FT MOD\_RES 22 22 PYROLIDONE CARBOXYLIC ACID.  
FT DISULFID 124 303 INTERCHAIN (GAMMA-DELTA) (POTENTIAL).  
FT CONFLICT 27 27 S -> E (IN REF. 2).  
FT CONFLICT 30 30 E -> S (IN REF. 2).  
SQ SEQUENCE 480 AA; 54625 MW; BCD8A83DD1AED93C CRC64;

Query Match 44.7%; Score 46; DB 1; Length 480;  
Best Local Similarity 57.9%; Pred. No. 11;  
Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 1 ORLKYAOEELNEVLP 19  
DB 458 ORLKGQDEM--RVLSPE 474  
|||||:|:| |||

RESULT 21  
ID HEXA\_BIADI STANDARD: PRT; 733 AA.  
AC Q17127;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hexamerin precursor.  
OS Haberus discoidalis (Tropical cockroach).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
OC Blaberoidea; Blaberidae; Blaberus.  
OX NCBI\_TaxID=6981;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fat body;  
RA Jamroz R.C., Beintema J.J., Stam W.T., Bradford J.J.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: LARVAL STORAGE PROTEIN (LSP) WHICH MAY SERVE AS A STORE  
CC OF AMINO ACIDS FOR SYNTHESIS OF ADULT PROTEINS (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOHXAMER (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).



AC P08266; 095027; Q04155; Q9VEM7;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6)  
DE (RNA polymerase II subunit 2).  
GN RPI140 OR CG3180.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephyroidae; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE OF 54-1176 FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=88011299; PubMed=3116266;  
RT Falkenburg D., Dworniczak B., Faust D.M., Bautz E.K.F.;  
RT "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit  
RT to the beta subunit of Escherichia coli RNA polymerase.";  
RT J. Mol. Biol. 195;929-937(1987).  
RL [2]  
RN SEQUENCE OF 1-69 FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=91276237; PubMed=1905256;  
RT Stitzler S., Oldenburg I., Peterson G., Bautz E.K.F.;  
RT "Analysis of the promoter region of the housekeeping gene DmRpl40 by  
RT sequence comparison of Drosophila melanogaster and Drosophila  
RT virilis.";  
RL Gene 100:155-162(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RT Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,  
RT Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RT George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RT Sultion G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RT Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RT Wen R.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RT Adair J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RT Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RT Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Boltslavsky S.,  
RT Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RT Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RT Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RT de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RT Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RT Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RT Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC (RNA)(N).  
CC -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14  
CC DIFFERENT POLYPEPTIDES. RNA POLYMERASE II CONSISTS OF 10 DIFFERENT  
CC SUBUNITS. THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA  
CC POLYMERASE II.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
CC III FOR 5S AND TRNA GENES.  
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X05709; CAA29180.2; -;  
DR EMBL: M62972; AAA28476.1; -;  
DR EMBL: AE003703; AAF55024.1; -;  
DR PIR: A27826; A27826.  
DR FLYBASE: FBgn0003276; RPI1140.  
DR InterPro: IPR001572; RNA\_POL\_B.  
DR Pfam: PF00562; RNA\_POL\_B.1.  
DR PROSITE: PS01166; RNA\_POL\_BETA.1.  
KW Transiferase; DNA-directed RNA polymerase; Transcription; Zinc;  
KW Zinc-finger; Metal-binding; Nuclear protein.  
FT ZN-FING 1121 1142 C4-TYPE (POTENTIAL).  
FT FT 72 72 A -> R (IN REF. 1).  
FT FT CONFLICT 666 667 ID -> MY (IN REF. 1).  
SQ SEQUENCE 1176 AA; 134042 MW; 224821B355BED7F0 CRC64;  
  
Query Match 46.6%; Score 48; DB 1; Length 1176;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 ORLKYAEELSNELVP 16  
Db 335 KRIRYARELLQKEMLP 350  
:::111:1 1-11  
  
RESULT 17  
RBP2\_LYCSES  
ID RBP2\_LYCSES STANDARD; PRT; 1191 AA.  
AC Q42877;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-directed RNA polymerase II 135 kDa polypeptide (EC 2.7.7.6)  
DE (RNA polymerase II subunit 2).  
GN RBP2.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OC Asteridae; euasterids I; Solanales; Solanales; Solanales;  
OX NCBI\_Taxid=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV GROSSE LISSE;  
RX MEDLINE=96178872; PubMed=8616257;  
RA Warlow D., Symons R.H.;  
RT "Sequence analysis of the second largest subunit of tomato RNA  
RT polymerase II.";  
RL Plant Mol. Biol. 30:337-342(1996).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.





RX MEDLINE-91311420; PubMed-1649901;  
RA Tsarev S.A., Emerson S.O., Balayan M.S., Ticehurst J.R.,  
RT Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
RT structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
[2]  
RN SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE-89232168; PubMed-2541023;  
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
RT and simian hepatitis A viruses.";  
RL FEBS Lett. 247:425-428(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: D00924; BAA00766.1; -  
DR EMBL: X15461; CAA33490.1; -  
DR PIR: A30470; GNNTSA.  
DR PIR: S04885; S04885.  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferrase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease;  
FT CHAIN 1 27  
FT CHAIN 28 249  
FT CHAIN 250 495  
FT CHAIN 496 795  
FT CHAIN 796 984  
FT CHAIN 985 1091  
FT CHAIN 1092 1426  
FT CHAIN 1427 1498  
FT CHAIN 1499 1521  
FT CHAIN 1522 1741  
FT CHAIN 1742 2230  
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;  
  
Query Match 94.2%; Score 97; DB 1; Length 2230;  
Best local Similarity 90.0%; Pred. No. 6,2e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 ORLKVAQEELSNEVLPK 20  
DB 814 QRLKVAQEELSNEVLPK 833  
||||| |||||||  
POLG-HPAVT STANDARD; PRT; 839 AA.  
AC P31788;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core protein  
DE P2A] (Fragment).  
OS Simian hepatitis A virus (strain CY-145).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.  
OX NCBI\_TaxID=31707;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91311421; PubMed-1649902;  
RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
RT "Sequence analysis of a new hepatitis A virus naturally infecting  
RT cynomolgus macaques (Macaca fascicularis).";  
RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: M59286; AAA5473.1; -  
DR PIR: J01180; GNNTS2.  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 ?  
FT CHAIN ? >839  
FT NON\_TER 839 839  
SQ SEQUENCE 839 AA; 93825 MW; 2CACCA4BD1E192DBC CRC64;  
  
Query Match 91.3%; Score 94; DB 1; Length 839;  
Best local Similarity 85.0%; Pred. No. 6,1e-07;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 ORLKVAQEELSNEVLPK 20  
DB 809 QREKVAQEELSNEVLPK 828  
||||| |||||||  
RBP2-CAEEL STANDARD; PRT; 1193 AA.  
AC 010578;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DNA-directed RNA polymerase II second largest subunit (EC 2.7.7.6)  
DE (RNA polymerase I subunit 2).  
GN C26E6.4  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Fulton L.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE OF 200-1058 FROM N.A.  
RP MEDLINE-95041334; PubMed-7953533;  
RX MEDLINE-95041334; PubMed-7953533;  
RA Sidow A., Thomas W.K.;  
RT "A molecular evolutionary framework for eukaryotic model organisms.";  
RT Curr. Biol. 4:596-603(1994).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC [RNA](N).

16-OCT-2001 (Rel. 40, last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC Hepatitis A virus;  
XX NCBI\_TaxID=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85190549; PubMed=2986127;  
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Meriweather J van Nest G., Dana D.,  
RT "Primary structure and gene organization of human hepatitis A virus,"  
Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: K02990; AAA45472.1; -  
DR PIR: A03903; GNNYHR.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polypeptide: Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1076  
FT CHAIN 1077 1422  
FT CHAIN 1423 1484  
FT CHAIN 1485 1507  
FT CHAIN 1508 1678  
FT CHAIN 1679 2227  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354BACD2799C CRC64;  
  
Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 7.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ORLKYAOEELSNEVLPPEPK 20  
DB 810 ORLKYAOEELSNEVLPPEPK 829  
POLG\_HPAVM STANDARD; PRT; 2227 AA.  
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MBH).

Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC Hepatitis A virus;  
XX NCBI\_TaxID=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wessel T., Klein R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
virus (isolate MBH).";  
RL Virus Res. 8:153-171(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M20273; AAA45474.1; -  
DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polypeptide: Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
  
Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 7.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ORLKYAOEELSNEVLPPEPK 20  
DB 810 ORLKYAOEELSNEVLPPEPK 829  
POLG\_HPAVS STANDARD; PRT; 2230 AA.  
AC P14553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC Hepatitis A virus;  
XX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.



```

CC -----
DR EMBL: M59808; AAA45467.1; -
DR MEROPS: C03.005; -
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol_1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Polypol: Coat protein; Core protein; Transferase;
  RNA-directed RNA polymerase; Thiol protease.
KW CHAIN 1 23 COAT PROTEIN VP4 (P1A).
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).
FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1495 CORE PROTEIN P3A.
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2226 AA; 251292 MW; 24964A6396C8DB CRC64;

```

```

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ORLKYAEELSNVLPK 20
    |||||
Db 810 ORLKYAEELSNVLPK 829

```

```

RESULT 7
POLG_HPAVL STANDARD; PRT; 2227 AA.
AC POLG_HPAVL 081082;
DT 01-AUG-1988 (Rel. 08; Created)
DT 01-AUG-1988 (Rel. 08; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
  P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
  P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain HM-175).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12098;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-WILD TYPE;
RA MEDLINE=87061253; PubMed=3023706;
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,
  Baroudy B.M.;
RA Baroudy B.M.;
RT *Complete nucleotide sequence of wild-type hepatitis A virus;
RT comparison with different strains of hepatitis A virus and other
  picornaviruses";
RL J. Virol. 61:50-59(1987).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-ATTENUATED;
RA MEDLINE=87175701; PubMed=3031686;
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Deemer R.J., Feinstone S.M.,
  Purcell R.H.;
RA *Complete nucleotide sequence of an attenuated hepatitis A virus;
RT comparison with wild-type virus";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).
RN [3]
RN SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.
RA MEDLINE=85166289; PubMed=2984684;
RA Baroudy B.M., Ticehurst J.R., Mele T.A., Maizel J.V. Jr.,
  Purcell R.H., Feinstone S.M.;
RA *Sequence analysis of hepatitis A virus cDNA coding for capsid
  proteins and RNA polymerase";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).

```

```

CC -1 SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1 PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
CC -1 SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1 CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT
CC SHOWN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

DR EMBL: M41114; AAA45475.1; -
DR EMBL: M41707; AAA45465.1; -
DR EMBL: M16632; AAA45471.1; -
DR PIR: A25981; GNNTMK.
DR PIR: A25914; GNNTMK.
DR PIR: A03905; A03905.
DR MEROPS: C03.005; -
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol_1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polypeptin: Coat protein; Core protein; Transferase;
  RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).
FT CHAIN 837 980 CORE PROTEIN P2A.
FT CHAIN 981 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.
FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.
FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.
FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.
FT CHAIN 77 77 R (IN ATTENUATED STRAIN).
FT VARIANT 764 764 N -> V (IN ATTENUATED STRAIN).
FT VARIANT 821 821 E -> S (IN ATTENUATED STRAIN).
FT VARIANT 1052 1052 N -> V (IN ATTENUATED STRAIN).
FT VARIANT 1052 1052 A -> V (IN ATTENUATED STRAIN).
FT VARIANT 1062 1062 G -> A (IN ATTENUATED STRAIN).
FT VARIANT 1118 1118 K -> M (IN ATTENUATED STRAIN).
FT VARIANT 1151 1151 E -> S (IN ATTENUATED STRAIN).
FT VARIANT 1163 1163 F -> S (IN ATTENUATED STRAIN).
FT VARIANT 1277 1277 V -> I (IN ATTENUATED STRAIN).
FT VARIANT 1500 1500 H -> Y (IN ATTENUATED STRAIN).
FT VARIANT 1805 1805 D -> N (IN ATTENUATED STRAIN).
FT VARIANT 1930 1930 S -> T (IN ATTENUATED STRAIN).
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AE8740A6 CRC64;

```

```

Query Match 100.0%; Score 103; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 ORLKYAEELSNVLPK 20
    |||||
Db 810 ORLKYAEELSNVLPK 829

```

```

RESULT 8
POLG_HPAVL STANDARD; PRT; 2227 AA.
ID POLG_HPAVL 081082;
AC P06441;
DT 01-JAN-1988 (Rel. 06; Created)
DT 01-JAN-1988 (Rel. 06; Last sequence update)

```

```
CC -1- P1M: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M59810; AAA45468.1; -.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
CC DR Pfam: PF00910; RNA_helicase; 1.
CC KM Polypeptin: Coat protein; Core protein; Transferase;
CC KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
CC FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
CC FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
CC FT CHAIN 492 794 COAT PROTEIN VP1 (PID).
CC FT CHAIN 795 900 CORE PROTEIN P2A.
CC FT CHAIN 901 1087 CORE PROTEIN P2B.
CC FT CHAIN 1088 1422 CORE PROTEIN P2C.
CC FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
CC FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
CC FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
CC FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
CC SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPEPK 20
ID POLG_HPAV4 STANDARD; PRT; 2226 AA.
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptin [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (RC 2.7.7.48)].
OS Hepatitis A virus (strain 43c).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91162758; PubMed-1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- P1M: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M59809; AAA45469.1; -.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
CC DR Pfam: PF00910; RNA_helicase; 1.
CC KM Polypeptin: Coat protein; Core protein; Transferase;
CC KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
CC FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
CC FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
CC FT CHAIN 492 794 COAT PROTEIN VP1 (PID).
CC FT CHAIN 795 900 CORE PROTEIN P2A.
CC FT CHAIN 901 1087 CORE PROTEIN P2B.
CC FT CHAIN 1088 1422 CORE PROTEIN P2C.
CC FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
CC FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
CC FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
CC FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
CC SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPEPK 20
ID POLG_HPAV8 STANDARD; PRT; 2226 AA.
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptin [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (RC 2.7.7.48)].
OS Hepatitis A virus (strain 18f).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91162758; PubMed-1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- P1M: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

SO SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 8.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLP PPK 20  
DB 314 ORLKYAOEELSNEVLP PPK 333

RESULT 2  
POLG\_HPAVG STANDARD; PRT; 808 AA.

AC 002381;  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
OS Hepatitis A virus (strain GA76).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=31706;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92260183; PubMed=1316423;  
RA Khanna B., Speilbring J.E., Innis B.L., Robertson B.H.;  
RT "Characterization of a genetic variant of human hepatitis A virus."  
RL J. Med. Virol. 36:118-124(1992).  
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
-1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M66695; AAA5477.1; -  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 1  
FT NON\_TER 1 1  
FT CHAIN 1 2  
FT CHAIN 3 223 COAT PROTEIN VP4 (P1A).  
FT CHAIN 224 470 COAT PROTEIN VP2 (P1B).  
FT CHAIN 471 770 COAT PROTEIN VP3 (P1C).  
FT CHAIN 771 >808 COAT PROTEIN VP1 (P1D).  
FT CHAIN 808 808 CORE PROTEIN P2A.  
FT NON\_TER 808 808  
SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 808;  
Best Local Similarity 100.0%; Pred. No. 2.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLP PPK 20  
DB 789 ORLKYAOEELSNEVLP PPK 808

RESULT 3  
POLG\_HPAVG STANDARD; PRT; 852 AA.

AC P06442; Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core protein

DE P2A] (Fragment).  
OS Hepatitis A virus (strain CR326).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85185648; PubMed=2985793;  
RA Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
RA Young A., Mitra S.W.;  
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA."  
RL J. Virol. 54:247-255(1985).  
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
-1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
-----  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M10033; AAA5470.1; -  
DR PIR; A03904; GNNYHA.  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245 COAT PROTEIN VP4 (P1A).  
FT CHAIN 246 491 COAT PROTEIN VP2 (P1B).  
FT CHAIN 492 836 COAT PROTEIN VP3 (P1C).  
FT CHAIN 837 >852 COAT PROTEIN VP1 (P1D).  
FT CHAIN 852 852 CORE PROTEIN P2A.  
FT NON\_TER 852 852  
SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;

Query Match 100.0%; Score 103; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 2.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLP PPK 20  
DB 810 ORLKYAOEELSNEVLP PPK 829

RESULT 4  
POLG\_HPAV2 STANDARD; PRT; 2226 AA.

AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,  
RA Cromean T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination."  
RL J. Virol. 65:2056-2065(1991).  
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:10:06 ; Search time 59.43 Seconds  
(without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432a-41  
Perfect score: 103  
Sequence: 1 ORLKYAQLNELNEVLPERRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	341	1	P13672 hepatitis a
2	103	100.0	808	1	P02381 hepatitis a
3	103	100.0	852	1	P06442 hepatitis a
4	103	100.0	2226	1	P26580 hepatitis a
5	103	100.0	2226	1	P26581 hepatitis a
6	103	100.0	2226	1	P26582 hepatitis a
7	103	100.0	2227	1	P08617 hepatitis a
8	103	100.0	2227	1	P06441 hepatitis a
9	103	100.0	2227	1	P13901 hepatitis a
10	97	94.2	2230	1	P14553 simian hepa
11	94	91.3	839	1	P11788 simian hepa
12	52	50.5	1193	1	P10578 caenorhabdit
13	49.5	48.1	443	1	P09077 vibrio chol
14	49	47.6	1174	1	P30876 homo sapien
15	48	46.6	636	1	P15025 homo sapien
16	48	46.6	1176	1	P08286 drosophila
17	48	46.6	1191	1	P042877 lycopersico
18	47	45.6	592	1	P15891 saccharomyc
19	46	44.7	338	1	P001664 homo sapien
20	46	44.7	480	1	P13744 cucurbita m
21	45	43.7	733	1	P17177 diabetus di
22	44	42.7	303	1	P012326 saccharomyc
23	44	42.7	510	1	P06464 anthrhamio
24	44	42.7	914	1	P54784 saccharomyc
25	43.5	42.2	3678	1	P11531 mus musculu
26	43.5	42.2	3680	1	P09752 canis famli
27	43.5	42.2	3685	1	P11532 homo sapien
28	43	41.7	268	1	P38740 saccharomyc
29	43	41.7	574	1	P03461 influenza b
30	43	41.7	576	1	P09766 influenza b
31	43	41.7	578	1	P09765 influenza b
32	43	41.7	578	1	P09767 influenza b
33	43	41.7	583	1	P10757 influenza b

34	43	41.7	583	1	HEMA_INBR	P03464 influenza b
35	43	41.7	583	1	HEMA_INBSI	P03463 influenza b
36	43	41.7	585	1	HEMA_INBSE	P17504 influenza b
37	43	41.7	585	1	HEMA_INBVK	P22092 influenza b
38	43	41.7	1146	1	AS10_YEAST	P48361 saccharomyc
39	43	41.7	1210	1	RPB2_SCHPO	P02061 schizosacch
40	42.5	41.3	2278	1	FAB1_YEAST	P34756 saccharomyc
41	42	40.8	214	1	DEVS_MYXXA	P07766 myxococcus
42	42	40.8	276	1	YJIC_ECOLI	P39374 escherichia
43	42	40.8	445	1	ML64_HUMAN	P14849 homo sapien
44	42	40.8	554	1	PEX2_PODAN	P51021 podospora a
45	42	40.8	1005	1	DPOL_VARY	P33793 varicella vir
46	42	40.8	1006	1	DPOL_VACCC	P20509 vaccinia vi
47	42	40.8	1006	1	DPOL_VACCV	P06856 vaccinia vi
48	42	40.8	1270	1	DDX9_HUMAN	P008211 homo sapien
49	41.5	40.3	498	1	PDM2_DROME	P31369 drosophila
50	41	39.8	181	1	ATPE_STRLI	P50013 streptomyce
51	41	39.8	258	1	RT15_MOUSE	P09d71 mus musculu
52	41	39.8	349	1	F161_RHIME	P09exv4 rhizobium m
53	41	39.8	349	1	F162_RHIME	P56886 rhizobium m
54	41	39.8	500	1	CISA_BACSU	P17867 bacillus su
55	41	39.8	512	1	F101_MOUSE	P41438 mus musculu
56	41	39.8	519	1	TRPE_SERMA	P00897 serrata ma
57	41	39.8	728	1	TP38_YEAST	P09289 saccharomyc
58	41	39.8	816	1	ATX1_HUMAN	P54253 homo sapien
59	41	39.8	1966	1	CCAF_HUMAN	P06840 homo sapien
60	40.5	39.3	327	1	HUNB_MANSE	P25514 manduca sex

## ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	341 AA.
1	POLG_HPAV1	01-JAN-1990 (Rel. 13, Created)			
2	P13672	01-JAN-1990 (Rel. 13, Last sequence update)			
3	DT	16-OCT-2001 (Rel. 40, Last annotation update)			
4	DE	Genome polypeptide [Contains: Coat proteins VP1 TO VP3; Core protein			
5	DE	2PA1 (Fragment).			
6	OS	Hepatitis A virus (strain LCD-1).			
7	OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;			
8	OC	Hepatovirus.			
9	OX	NCBI_TaxID=12093;			
10	RN	[1]			
11	RP	SEQUENCE FROM N.A.			
12	RX	MEDLINE=9263805; PubMed=2542903;			
13	RA	Andonov A.P., Lau P., Chaudhary R.;			
14	RT	Nucleotide sequence of the VP1 gene from a Chinese strain of			
15	RL	hepatitis A virus (HAV)."			
16	RL	Nucleic Acids Res. 17:3594-3594(1989).			
17	CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,			
18	CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,			
19	CC	VP3, AND VP4.			
20	CC	-1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.			
21	CC	-----			
22	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
23	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
24	CC	the European Bioinformatics Institute. There are no restrictions on its			
25	CC	use by non-profit institutions as long as its content is in no way			
26	CC	modified and this statement is not removed. Usage by and for commercial			
27	CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
28	CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
29	CC	-----			
30	DR	EMBL: X14666; CA32794.1; -.			
31	DR	PIR: S04137; S04137.			
32	KW	Polypeptide; Coat protein; Core protein.			
33	FT	NON_TER	1		
34	FT	CHAIN	<1	1	COAT PROTEIN VP3 (1C).
35	FT	CHAIN	2	340	COAT PROTEIN VP1 (1D).
36	FT	CHAIN	341	>341	CORE PROTEIN P2A.
37	FT	NON_TER	341		

**THIS PAGE BLANK (USPTO)**

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/035,098
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/454,267
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: P-41,264
; REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 295 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-035-098-7

```

```

Query Match          39.8%; Score 41; DB 4; Length 295;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      2 RLKTAQELSLNEVLPFR 19
      :|||:::|||
Db      100 KLKAAQEVAASSLPLVPR 117

```

Search completed: June 16, 2002, 00:03:15  
 Job time: 8529 sec

TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2543  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 445 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-691-814B-6

Query Match 40.8%; Score 42; DB 2; Length 445;  
Best Local Similarity 53.8%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 8 EELSNEVLPPPK 20  
: ||| : ||| |  
Db 30 QSLSSHLLPPPK 42

RESULT 23  
US-08-454-267-7  
; Sequence 7, Application US/08454267  
; Patent No. 5843739  
; GENERAL INFORMATION:  
; APPLICANT: SLABAS, ANTONI R.  
; APPLICANT: BROWN, ADRIAN P.  
; TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,267  
; FILING DATE: 08-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB93/02528  
; FILING DATE: 10-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, GRANT E.  
; REGISTRATION NUMBER: P-41,264  
; REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-454-267-7

Query Match 39.8%; Score 41; DB 2; Length 295;  
Best Local Similarity 50.0%; Pred. No. 50;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 RLKYOELSNEVLPPPK 19  
: || ||| : || |||  
Db 100 KLKAAOEYAASSSELPPVR 117

RESULT 24  
US-08-941-319-7  
; Sequence 7, Application US/08941319  
; Patent No. 5945323  
; GENERAL INFORMATION:  
; APPLICANT: SLABAS, ANTONI R.  
; APPLICANT: BROWN, ADRIAN P.  
; TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/941,319  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/454,267  
; FILING DATE: 08-JUN-1995  
; APPLICATION NUMBER: PCT/GB93/02528  
; FILING DATE: 10-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, GRANT E.  
; REGISTRATION NUMBER: P-41,264  
; REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-941-319-7

Query Match 39.8%; Score 41; DB 2; Length 295;  
Best Local Similarity 50.0%; Pred. No. 50;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 RLKYOELSNEVLPPPK 19  
: || ||| : || |||  
Db 100 KLKAAOEYAASSSELPPVR 117

RESULT 25  
US-09-035-098-7  
; Sequence 7, Application US/09035098  
; Patent No. 6194640  
; GENERAL INFORMATION:  
; APPLICANT: SLABAS, ANTONI R.  
; APPLICANT: BROWN, ADRIAN P.  
; TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Shanghai/4/94 rHA  
FEATURE:  
NAME/KEY: AcNPV 61k protein signal peptide  
LOCATION: 1 to 18  
FEATURE:  
NAME/KEY: mature rHA  
LOCATION: 19 to 574  
US-08-453-848-17

Query Match 41.7%; Score 43; DB 2; Length 592;  
Best Local Similarity 60.0%; Pred. No. 52;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVL 15  
||| | : || || : |  
DB 433 ORLSCAMDELHNEIL 447

RESULT 21  
US-09-169-027-17  
Sequence 17, Application US/09169027  
Patent No. 6245532  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volvovitz, Franklin  
APPLICANT: Wilkinson, Bethanie Eident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanway  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/169,027  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993  
ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Shanghai/4/94 rHA  
FEATURE:  
NAME/KEY: AcNPV 61k protein signal peptide  
LOCATION: 1 to 18  
FEATURE:  
NAME/KEY: mature rHA  
LOCATION: 19 to 574  
US-09-169-027-17

Query Match 41.7%; Score 43; DB 4; Length 592;  
Best Local Similarity 60.0%; Pred. No. 52;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVL 15  
||| | : || || : |  
DB 433 ORLSCAMDELHNEIL 447

RESULT 22  
US-08-691-814B-6  
Sequence 6, Application US/08691814B  
Patent No. 5981218  
GENERAL INFORMATION:  
APPLICANT: Rio, Marie-Christine  
APPLICANT: Tomasetto, Catherine  
APPLICANT: Bassett, Paul  
APPLICANT: Byrne, Jennifer  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful  
TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave, NW, Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/691,814B  
FILING DATE: 31-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,183  
FILING DATE: 09-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1383, 0090001  
TELECOMMUNICATION INFORMATION:



ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Netherlands/13/94 rHA  
FEATURE:  
NAME/KEY: ACPV 61k protein signal sequence  
LOCATION: 1 to 18  
NAME/KEY: mature rHA  
LOCATION: 19 to 571  
US-08-453-848-13

Query Match 41.7% Score 43; DB 2; Length 589;  
Best Local Similarity 60.0%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNLYL 15  
||| | :|| ||:|  
Db 430 ORLSGAMDELHNEIL 444

RESULT 19  
US-09-169-027-13  
Sequence 13, Application US/09169027  
Patent No. 6245532  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volkovitz, Franklin  
APPLICANT: Wilkinson, Bethanie Eident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanway  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/169,027  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Netherlands/13/94 rHA  
FEATURE:  
NAME/KEY: ACPV 61k protein signal sequence  
LOCATION: 1 to 18  
NAME/KEY: mature rHA  
LOCATION: 19 to 571  
US-09-169-027-13

Query Match 41.7% Score 43; DB 4; Length 589;  
Best Local Similarity 60.0%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNLYL 15  
||| | :|| ||:|  
Db 430 ORLSGAMDELHNEIL 444

RESULT 20  
US-08-453-848-17  
Sequence 17, Application US/08453848  
Patent No. 5838368  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volkovitz, Franklin  
APPLICANT: Wilkinson, Bethanie Eident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanway  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 586 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Harbin/7/94 rHA  
FEATURE:  
NAME/KEY: HA signal peptide  
LOCATION: 1 to 17  
NAME/KEY: mature rHA  
LOCATION: 18 to 569  
US-08-453-848-19

Query Match 41.7%; Score 43; DB 2; Length 586;  
Best Local Similarity 60.0%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ORLKYAOEELSNEVL 15  
||| | : || || : |  
DB 427 ORLSGAMDELHNEIL 441

RESULT 17  
US-09-169-027-19  
Sequence 19, Application US/09169027  
Patent No. 6245532  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volvovitz, Franklin  
APPLICANT: Wilkenson, Bethanie Elident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanley  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/169,027  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
APPLICATION NUMBER: 08/120,607

FILING DATE: 13-SEPT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 586 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Harbin/7/94 rHA  
FEATURE:  
NAME/KEY: HA signal peptide  
LOCATION: 1 to 17  
NAME/KEY: mature rHA  
LOCATION: 18 to 569  
US-09-169-027-19

Query Match 41.7%; Score 43; DB 4; Length 586;  
Best Local Similarity 60.0%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ORLKYAOEELSNEVL 15  
||| | : || || : |  
DB 427 ORLSGAMDELHNEIL 441

RESULT 18  
US-08-453-848-13  
Sequence 13, Application US/08453848  
Patent No. 5858368  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volvovitz, Franklin  
APPLICANT: Wilkenson, Bethanie Elident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanley  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993  
CLASSIFICATION: 435



TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-241-853-15

Query Match 42.2%; Score 43.5; DB 1; Length 113;  
Best Local Similarity 55.6%; Pred. No. 7.2;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
Qy 3 LKVAQELSNVLP 20  
Db 7 VKHAQELPP---PPPK 21

RESULT 12  
US-08-850-917-15  
Sequence 15, Application US/08850917  
Patent No. 5854045  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanafusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850.917  
FILING DATE: 02-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/241.853  
FILING DATE: 12-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-850-917-15

Query Match 42.2%; Score 43.5; DB 2; Length 113;

Best Local Similarity 55.6%; Pred. No. 7.2;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
Qy 3 LKVAQELSNVLP 20  
Db 7 VKHAQELPP---PPPK 21

RESULT 13  
US-09-308-003-12  
Sequence 12, Application US/09308003  
Patent No. 6326170  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,  
TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES  
FILE REFERENCE: GMI0093  
CURRENT APPLICATION NUMBER: US/09/308.003  
CURRENT FILING DATE: 1999-05-10  
EARLIER APPLICATION NUMBER: 60/058,710  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 390  
TYPE: PRN  
ORGANISM: Staphylococcus aureus  
US-09-308-003-12

Query Match 41.7%; Score 43; DB 4; Length 390;  
Best Local Similarity 57.1%; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 LKVAQELSNVLP 16  
Db 368 IKYALKEASNESIP 381

RESULT 14  
US-08-453-848-11  
Sequence 11, Application US/08453848  
Patent No. 5858368  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Voliovitz, Franklin  
APPLICANT: Wilkinson, Bethanie Eident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanley  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453.848  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/120,607



## HAVING A CHIMERIC 2C PROTEIN

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN &amp; FINNEGAN, L.L.P.

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MICROSOFT WORD 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/171,387

FILING DATE: 24-Mar-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/06506

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US60/015,642

FILING DATE: 19-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: William S. Feller

REGISTRATION NUMBER: 26,728

REFERENCE/DOCKET NUMBER: 2026-4229US1

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:

LENGTH: 2227 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 2

US-09-171-387-2

Query Match 100.0%; Score 103; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 7.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLPPRK 20

DB 810 ORLKYAOEELSNEVLPPRK 829

## RESULT 7

US-08-475-886-4

Sequence 4, Application US/08475886A

Patent No. 6113912

GENERAL INFORMATION:

APPLICANT: FUNKHOUSER, ANN W

APPLICANT: EMERSON, SUZANNE U

APPLICANT: PURCELL, ROBERT H

APPLICANT: D'HONDT, ERIC

TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES

FILE REFERENCE: 20264262US2

CURRENT APPLICATION NUMBER: US/08/475,886A

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 07/947,338

EARLIER FILING DATE: 1992-09-18

EARLIER APPLICATION NUMBER: 08/397,232

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 2227

TYPE: PRT

ORGANISM: Attenuated HAV (Pass 35), strain HM-175

US-08-475-886-4

Query Match 95.1%; Score 98; DB 3; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 4.8e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLPPRK 20

DB 810 ORLKYAOEELSNEVLPPRK 829

## RESULT 8

US-08-087-016-2

Sequence 2, Application US/08087016

Patent No. 5430135

GENERAL INFORMATION:

APPLICANT: NAINAN, OMANA V.

APPLICANT: MARGOLIS, HAROLD S.

APPLICANT: ROBERTSON, BETTY H.

APPLICANT: BRINTON, MARGO H.

APPLICANT: EBERT, JAMES W.

TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN, DARBY &amp; CUSHMAN

STREET: 1615 L Street N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/087,016

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,828

FILING DATE: 03-APR-1991

ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26,581

REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL

TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 839 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-087-016-2

Query Match 91.3%; Score 94; DB 1; Length 839;  
Best Local Similarity 85.0%; Pred. No. 7.2e-07;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLPPRK 20

DB 809 ORLKYAOEELSNEVLPPRK 828

## RESULT 9

US-08-484-105-2

Sequence 2, Application US/08484105

Patent No. 5589341

GENERAL INFORMATION:

APPLICANT: STILLMAN, BRUCE

```

; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

```

```

Query Match          100.0%; Score 103; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ORLKVAQELSNVLPPrK 20
    |||
Db 810 ORLKVAQELSNVLPPrK 829

```

```

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

```

```

Query Match          100.0%; Score 103; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ORLKVAQELSNVLPPrK 20
    |||
Db 810 ORLKVAQELSNVLPPrK 829

```

```

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W

```

```

; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

```

```

Query Match          100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ORLKVAQELSNVLPPrK 20
    |||
Db 810 ORLKVAQELSNVLPPrK 829

```

```

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

```

```

Query Match          100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ORLKVAQELSNVLPPrK 20
    |||
Db 810 ORLKVAQELSNVLPPrK 829

```

```

RESULT 6
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHAUDHURI, GOPA;
; PURCELL, ROBERT, H.;
; TITLE OF INVENTION: SIMIAN-HUMAN HAV

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:14 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-41

Perfect score: 103

Sequence: 1 ORLKYAOEELSNEVLPPRK 20

Scoring table: BLOSUM62

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6D.COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	1091	6	5516630-2
2	103	100.0	2227	3	US-08-475-886-2
3	103	100.0	2227	3	US-08-475-886-6
4	103	100.0	2227	4	US-08-397-232-2
5	103	100.0	2227	4	US-08-397-232-4
6	103	100.0	2227	4	US-09-171-387-2
7	98	95.1	2227	3	US-08-475-886-4
8	94	91.3	839	1	US-08-087-016-2
9	44	42.7	914	1	US-08-484-105-2
10	44	42.7	914	1	US-08-484-106-2
11	43.5	42.2	113	1	US-08-241-853-15
12	43.5	42.2	113	1	US-08-850-917-15
13	43	41.7	390	4	US-09-308-003-12
14	43	41.7	585	4	US-08-453-848-11
15	43	41.7	585	4	US-09-169-027-11
16	43	41.7	586	4	US-08-453-848-19
17	43	41.7	586	4	US-09-169-027-19
18	43	41.7	589	2	US-08-453-848-13
19	43	41.7	589	2	US-09-169-027-13
20	43	41.7	592	4	US-08-453-848-17
21	43	41.7	592	4	US-09-169-027-17
22	42	40.8	445	2	US-08-691-814B-6
23	41	39.8	295	2	US-08-454-267-7
24	41	39.8	295	2	US-08-941-319-7
25	41	39.8	295	2	US-09-035-098-7
26	41	39.8	460	2	US-08-933-821-17
27	41	39.8	460	3	US-08-934-494-6

28	41	39.8	460	3	US-08-960-507-17	Sequence 17, Appl
29	41	39.8	460	3	US-09-143-068-6	Sequence 6, Appl1
30	41	39.8	460	4	US-09-143-707-6	Sequence 6, Appl1
31	41	39.8	460	4	US-09-202-089-6	Sequence 6, Appl1
32	41	39.8	460	4	US-09-136-828-17	Sequence 17, Appl
33	41	39.8	502	1	US-08-484-840-3	Sequence 3, Appl1
34	41	39.8	502	1	US-08-483-094-3	Sequence 3, Appl1
35	41	39.8	816	2	US-08-267-803B-9	Sequence 9, Appl1
36	41	39.8	816	2	US-09-041-886-17	Sequence 9, Appl1
37	41	39.8	1719	2	US-08-459-568-4	Sequence 4, Appl1
38	41	39.8	1719	2	US-08-399-411-4	Sequence 4, Appl1
39	41	39.8	1719	3	US-08-516-859A-4	Sequence 4, Appl1
40	39	37.9	553	1	US-08-328-342-10	Sequence 10, Appl
41	39	37.9	1464	4	US-08-891-640-2	Sequence 2, Appl1
42	38	36.9	300	2	US-08-946-528-7	Sequence 7, Appl1
43	38	36.9	345	3	US-08-843-993-1	Sequence 1, Appl1
44	38	36.9	345	3	US-09-059-520A-1	Sequence 1, Appl1
45	38	36.9	345	3	US-09-334-275-1	Sequence 1, Appl1
46	38	36.9	524	4	US-09-230-388-1	Sequence 1, Appl1
47	38	36.9	580	4	US-09-188-930-307	Sequence 307, App
48	38	36.9	1025	5	PCT-US95-04567-2	Sequence 2, Appl1
49	38	36.9	2802	4	US-09-542-331-1	Sequence 1, Appl1
50	37	35.9	118	2	US-08-417-174-2	Sequence 2, Appl1
51	37	35.9	118	2	US-08-231-565A-2	Sequence 2, Appl1
52	37	35.9	118	2	US-09-007-961-2	Sequence 2, Appl1
53	37	35.9	118	4	US-09-267-439-2	Sequence 2, Appl1
54	37	35.9	201	4	US-09-311-311C-21	Sequence 21, Appl
55	37	35.9	218	4	US-09-249-180-6	Sequence 6, Appl1
56	37	35.9	247	2	US-08-463-911-2	Sequence 2, Appl1
57	37	35.9	247	4	US-09-140-804-8	Sequence 8, Appl1
58	37	35.9	247	4	US-09-118-408-3	Sequence 3, Appl1
59	37	35.9	344	1	US-08-843-993-3	Sequence 3, Appl1
60	37	35.9	344	3	US-09-059-520A-3	Sequence 3, Appl1

## ALIGNMENTS

```
RESULT 1
5516630-2
Patent No. 5516630
APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,
STEPHEN W.; PORCELL, ROBERT H.; RACANIELLO, VINCENT R.;
BAROUDY, BAHIGE M.
TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/788,262
FILING DATE: 06-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 256,135
FILING DATE: 06-OCT-1988
APPLICATION NUMBER: 654,942
FILING DATE: 27-SEP-1984
APPLICATION NUMBER: 537,911
FILING DATE: 30-SEP-1983
SEQ ID NO: 2
LENGTH: 1091
5516630-2

Query Match      100.0%; Score 103; DB 6; Length 1091;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 ORLKYAOEELSNEVLPPRK 20
Db 1047 ORLKYAOEELSNEVLPPRK 1066

RESULT 2
US-08-475-886-2
; Sequence 2, Application US/08475886A
```



**THIS PAGE BLANK (USPTO)**



## RESULT 24

ID ABB31812 standard; Peptide: 263 AA.

AC ABB31812;

DT 01-FEB-2002 (first entry)

DE Peptide #4463 encoded by breast cell single exon nucleic acid probe.

KW Human; microarray; single exon probe; gene expression; breast; disease; cancer.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,

PT useful for measuring gene expression in sample derived from human

PS breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 14780; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting

CC the probes with a collection of detectably labelled nucleic acids

CC derived from mRNA of human breast, and then measuring the label

CC bound to each probe of the microarray. The probes are useful for

CC verifying the expression of regions of genomic DNA predicted to

CC encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene

CC expression analysis is useful for assessing the toxicity of chemical

CC agents on cells. The microarray of this invention presents a far greater

CC diversity of probes for measuring gene expression, with far less bias

CC than expressed sequence tag microarrays. The method is suitable for

CC rapid production of functional information from genomic sequence. The

CC present sequence is a peptide encoded by a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 263 AA;

SQ

Query Match 43.7%; Score 45; DB 22; Length 263;

Best Local Similarity 46.7%; Pred. No. 51;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 YAOEELSNEVLPPR 19

Db 174 yteqglynhvipkpr 188

## RESULT 25

ID ABB37043 standard; Peptide: 263 AA.

AC ABB37043;

DT 04-FEB-2002 (first entry)

DE Peptide #4549 encoded by human foetal liver single exon probe.

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver

PS Claim 27; SEQ ID NO 29678; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 263 AA;

SQ

Query Match 43.7%; Score 45; DB 22; Length 263;

Best Local Similarity 46.7%; Pred. No. 51;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 YAOEELSNEVLPPR 19

Db 174 yteqglynhvipkpr 188

Search completed: June 16, 2002, 00:01:43  
Job time: 12732 sec

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
CC Sequence 221 AA:  
SQ  
  
Query Match 45.6%; Score 47; DB 22; Length 221;  
Best Local Similarity 47.6%; Pred. No. 21;  
Matches 10; Conservative 4; Mismatches 3; Indels 4; Gaps 1;  
QY 1 ORLKYAQ----EELSNEVLPP 17  
:|||||: :|||: |||  
Db 128 rllkyqrtdvsdianeylpp 148  
  
RESULT 22  
AAW42923  
ID AAW42923 standard; peptide: 20 AA.  
XX  
AC AAW42923;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1316.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudiyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
CC Peptides AAW42923-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 799-818, and has a reactivity of 41.7% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal.  
SQ Sequence 20 AA:

Query Match 43.7%; Score 45; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ORLKYAQEE 9  
:|||||: |||||: |||  
Db 12 qrlkyaqee 20  
  
RESULT 23  
AAB69440  
ID AAB69440 standard; Peptide: 21 AA.  
XX  
AC AAB69440;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Synthetic HAV P2A peptide, SEQ ID NO: 40.  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
PN WO200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000WO-US19267.  
XX  
PR 15-JUL-1999; 99US-0144412.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudiyakov YE;  
XX  
DR WPI; 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines -  
XX  
PS Claim 13; Page 93; 130pp; English.  
XX  
CC The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
SQ Sequence 21 AA:  
QY 1 ORLKYAQEE 9  
:|||||: |||||: |||  
Db 12 qrlkyaqee 20

PD 19-DEC-1996.  
XX  
XX 06-JUN-1996; 96MO-US09122.  
XX  
XX 01-APR-1996; 96US-0630405.  
PR 07-JUN-1995; 95US-0487032.  
XX  
XX (ASTR ) ASTRA AB.  
XX  
XX Berglindh CF, Smith D, Møllgaard BL;  
XX  
XX WPI: 1997-052306/05.  
DR N-PSDB; AAT67877.  
XX  
XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
infection, and to detect Helicobacter  
XX  
XX Claim 61; Page 1050-51; 1481pp; English.  
XX  
XX This sequence represents a H. pylori cytoplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 35679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
XX  
XX Sequence 444 AA;  
SQ

Query Match 46.6%; Score 48; DB 18; Length 444;  
Best Local Similarity 75.0%; Pred. NO. 31;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 9 ELSNEVLPPRK 20  
| | | | | | | | | |  
DB 416 eadnee1pprk 427

RESULT 20  
ABB59692  
ID ABB59692 standard; Protein: 1176 AA.  
XX  
XX ABB59692;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 5868.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL03795.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 5868; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1176 AA;  
SQ

Query Match 46.6%; Score 48; DB 22; Length 1176;  
Best Local Similarity 56.2%; Pred. NO. 85;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 QRLKYAOEUSNEVLP 16  
: : : : : : : : : :  
DB 335 krikyakel1qkemp1 350

RESULT 21  
ABB68504  
ID ABB68504 standard; Protein: 221 AA.  
XX  
XX ABB68504;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 32304.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL12607.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 32304; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent

```

FT      Protein      1..245
FT      /label= VP0
FT      Protein      246..491
FT      /label= VP3
FT      Protein      492..791
FT      /label= VP1
FT      Protein      792
FT      /label= P2
FT      /note= "incomplete"
FT      Cleavage-site 245..246
FT      Cleavage-site 491..492
FT      Cleavage-site 791..792
FT      Active-site   315
FT      Active-site   593
XX      XX
XX      USN7678828-N.
XX      12-NOV-1991.
XX      PD
XX      03-APR-1991; 91US-0678828.
XX      PR
XX      03-APR-1991; 91US-0678828.
XX      PA
XX      (USSH ) US DEPT HEALTH & HUMAN.
XX      PI
XX      Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;
XX      DR
XX      WPI; 1991-376737/51.
XX      N-PSDB; AAO15180.
XX      PT
XX      Hepatitis A virus isolates and DNA - used to prepare vaccines for
XX      preventing hepatitis A virus infection.
XX      PS
XX      Disclosure: Fig 3; 23pp; English.
XX      CC
XX      The sequence was deduced from the nucleotide sequence obtd. by PCR
XX      amplification of cyno-HAV viral RNA obtd. from the stool of a
XX      cynomolgus monkey with serologically and histologically confirmed
XX      spontaneous hepatitis A. The sequence differs from the human HAV
XX      isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.
XX      USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val
XX      pair at the VP3-VP1 cleavage site in the human isolate is replaced
XX      by a Gln-Thr pair in the cyno-HAV. The other two cleavage sites are
XX      the same. Two residues have been identified as part of the immuno-
XX      dominant region (see feature table) and are different to those in
XX      the same position in human HAV. The protein and peptides derived
XX      from it can be used in the prepn. of vaccines for the prevention of
XX      HAV infection.
XX      CC
XX      See also AAR15056.
XX      CC
XX      (Note: Revised entry submitted to correct the patent number format of
XX      US Government-owned NTIS applications to prevent clashes with ongoing US
XX      granted patent numbers. For further information please visit the Derwent
XX      web site at www.derwent.com/dwpl/updates/ntis_us.html.)
XX      CC
XX      Sequence 839 AA;
SQ

```

Query Match 91.3%; Score 94; DB 12; Length 839;  
 Best Local Similarity 85.0%; Pred. No. 5,1e-06;  
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORLKTAQEELESLNEVLP 20  
 II IIII:IIIIII:IIIIII  
 DB 809 qrfkyareelsnelprrk 828

RESULT 18  
 ABB68507  
 ID ABB68507 standard; Protein; 150 AA.  
 XX  
 AC ABB68507;  
 XX  
 DT 26-MAR-2002 (first entry)

```

XX      XX
XX      DE
XX      Drosophila melanogaster polypeptide SEQ ID NO 32313.
XX      XX
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical.
XX      KW
XX      Drosophila melanogaster.
XX      OS
XX      WO200171042-A2.
XX      PN
XX      27-SEP-2001.
XX      PD
XX      23-MAR-2001; 2001WO-US09231.
XX      PE
XX      23-MAR-2000; 2000US-191637P.
XX      PR
XX      11-JUL-2000; 2000US-0614150.
XX      PA
XX      (PEKE ) PE CORP NY.
XX      PI
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX      DR
XX      WPI; 2001-656860/75.
XX      N-PSDB; ABL12610.
XX      PT
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
XX      PS
XX      Disclosure: SEQ ID NO 32313; 21pp + Sequence Listing; English.
XX      CC
XX      The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signalling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutics and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (ABL16176-ABL20511), expressed DNA
XX      sequences (ABL01840-ABL16175) and the encoded proteins
XX      CC
XX      (ABB57737-ABB72072).
XX      CC
XX      The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pcr_sequences.
XX      CC
XX      Sequence 150 AA;
SQ

```

Query Match 48.5%; Score 50; DB 22; Length 150;  
 Best Local Similarity 52.4%; Pred. No. 4.9;  
 Matches 11; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 1 ORLKTAQ---EELSNEVLP 17  
 :IIII I :::II III  
 DB 32 rllkygrdrvselaneylpp 52

RESULT 19  
 AAW20624  
 ID AAW20624 standard; Protein; 444 AA.  
 XX  
 AC AAW20624;  
 XX  
 DT 14-JUL-1997 (first entry)  
 DE  
 DE H. pylori cytoplasmic protein, 02ae11612orf25.  
 KW  
 Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO9640893-A1.  
 XX

CC hepatitis A in humans and other primates.  
 XX  
 SQ Sequence 2227 AA;

Query Match 100.0%; Score 103; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ORLKYAOEELSEVLPpPRK 20  
 ||||||||||||||||||  
 Db 810 qrlkyageelsnevlpPRK 829

RESULT 15  
 AAB18609  
 ID AAB18609 standard; Protein: 2227 AA.  
 XX  
 AC AAB18609;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KM HAV 4380.

XX Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PF 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93MO-US08610.

PR 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI: 2000-586464/55.

DR N-PSDB; AAA75478.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX Disclosure; Columns 93-104; 72pp; English.

XX The present sequence is derived from a live attenuated hepatitis A

CC virus (HAV) of the invention, designated HAV 4380. The sequence is

CC produced by modifying wild type HAV strain HM-174. The HAV of the

CC invention are adapted to growth in the human fibroblast-like cell

CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain

CC appropriate attenuation. It is useful as a live vaccine for prophylaxis

CC of hepatitis A in humans and other primates.

RESULT 16

AAB18608  
 ID AAB18608 standard; Protein: 2227 AA.  
 XX  
 AC AAB18608;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.

KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KM P-35 virus.

XX Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PF 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93MO-US08610.

PR 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI: 2000-586464/55.

DR N-PSDB; AAA75477.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX Disclosure; Columns 67-78; 72pp; English.

XX The present sequence is derived from passage 35 of a wild type

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is

CC designated P-35 virus. The sequence is modified to produce HAV which

CC are adapted to growth in the human fibroblast-like cell line MRC-5.

CC The HAV is able to propagate in MRC-5 cells and retain appropriate

CC attenuation. It is useful as a live vaccine for prophylaxis of

CC hepatitis A in humans and other primates.

XX Sequence 2227 AA;

QY 1 ORLKYAOEELSEVLPpPRK 20

Db 810 qrlkyageelsnevlpPRK 829

RESULT 17

AAR15629

ID AAR15629 standard; Protein: 839 AA.

AC AAR15629;

XX 17-DEC-2001 (updated)

DT 17-MAR-1992 (first entry)

XX Capsid region of cyno-HAV isolate CY-145.

DE Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.

XX Cynomolgus monkey hepatitis A virus, isolate CY-145.

OS Cynomolgus monkey hepatitis A virus, isolate CY-145.

XX key Location/Qualifiers





OY 1 ORLKYAEELSNEVLPPPRK 20  
|||||  
Db 1047 qrlkyageelsnevlppprk 1066

## RESULT 11

AAP60066  
ID AAP60066 standard; Protein: 2227 AA.

AC AAP60066;

DT 26-JUN-1991 (first entry)

DE Sequence of viral I434 polypeptide encoded by the complete  
nucleotide sequence of the HAV genome.

DG Diagnosis; vaccine; passive immunotherapy.

OS Hepatitis A virus.

Key Location/Qualifiers

FT 1..245 /label= P1.1A

FT 246..491 /label= 1B

FT 492..836 /label= 1C

FT 837..980 /label= P2.2A

FT 981..1076 /label= 2B

FT 1077..1422 /label= 2C

FT 1423..1484 /label= P3.3A

FT 1485..1507 /label= 3B

FT 1508..1678 /label= 3C

FT 1679..2227 /label= 3D

PN EPI99480-A.

PD 29-OCT-1986.

PF 03-APR-1986; 86EP-0302465.

PR 03-APR-1985; 85US-0719329.

PA (CHIR-) CHIRON CORP.

PI Dina D, Potter SJ, Vannest GA, Caput D;

DR WPI: 1986-286213/44.

DR N-PSDB: AAN60080.

PT Hepatitis A virus nucleotide sequence and polypeptide - and use  
in prodn. of vaccines and diagnostic probes

PS Claim 5; Fig 1; 18pp; English.

CC AAN60080 and oligonucleotide fragments are useful in detection of  
hepatitis A virus; transformed hosts may be used for expression of  
polypeptides and fragments useful in vaccines without risk of  
infection by the virus or in prodn. of particles which are capable  
of inducing immunocompetent B cells for passive immunotherapy. Pref.  
epitope is derived from AAs 445-657 or 792-848 of the HAV  
polypeptide sequence (AAP60066).

SO Sequence 2227 AA;

Query Match 100.0%; Score 103; DB 7; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 5.9e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 ORLKYAEELSNEVLPPPRK 20  
|||||  
Db 810 qrlkyageelsnevlppprk 829

## RESULT 12

AAR05697  
ID AAR05697 standard; protein: 2227 AA.

AC AAR05697;

DT 15-AUG-1990 (first entry)

DE Attenuated hepatitis A virus.

DG Hepatitis A virus; vaccine; attenuated.

OS Hepatitis A virus, strain HM-175.

Key Location/Qualifiers

FT 1..23 /label=VP4 = 1A

FT 24..245 /label=VP2 = 1B

FT 246..491 /label=VP3 = 1C

FT 492..791 /label=VP1 = 1D

FT 792..980 /label=2A

FT 981..1087 /label=2B

FT 1088..1422 /label=2C

FT 1423..1496 /label=3A

FT 1497..1519 /label=3B = VPg

FT 1520..1738 /label=3C

FT 1739..2227 /label=3D

PN US4894228-A.

PD 16-JAN-1990.

PF 12-JUL-1988; 88US-0217824.

PR 12-JUL-1988; 88US-0217824.

PR 12-JUL-1988; 88US-0652967.

PA (USSH ) US DEPT HEALTH & HUMAN.

PI Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;

PI Daemer RJ, Gust ID;

DR WPI: 1990-075557/10.

DR N-PSDB: AAO03512.

PT Vaccine against hepatitis A virus infection - comprises novel  
attenuated hepatitis A virus strain.

PS Claim 1; Fig 1; 18pp; English.

CC The attenuated HAV is useful for inducing protective immunity against  
HAV. This strain (pass 35) differs from the wild type HAV HM-175 by  
several nucleotide changes distributed throughout the genome, is  
attenuated for chimpanzees, elicits serum neutralising antibodies, and is

XX Sequence 993 AA:

Query Match 100.0%; Score 103; DB 6; Length 993;  
Best Local Similarity 100.0%; Pred. No. 2.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAQEELSNEVLPPPRK 20  
DB 946 grlkyageelsnevlppprk 965

# RESULT 9

AAW95559  
ID AAW95559 standard; Protein; 1077 AA.

XX AAW95559;

DT 28-APR-1999 (first entry)

XX A partial hepatitis A virus (HAV) protein.

XX Hepatitis A virus protein; HAV; P2 region;  
cell-culture-adapted HAV strain; infection; accelerated growth.

XX Hepatitis A virus.

XX US5849562-A.

XX 15-DEC-1998.

XX 06-JUN-1995; 9505-0468926.

XX 06-NOV-1991; 91US-0788262.

XX 30-SEP-1983; 83US-0537911.

XX 27-SEP-1984; 84US-0654942.

XX 06-OCT-1988; 88US-0256135.

XX 06-JUN-1995; 9505-0468926.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Emerson SU, Purcell RH;

XX WPI: 1999-094412/08.

XX N-PSDB; AAX01006.

XX Chimeric hepatitis A virus strains - with P2 region from

XX cell-culture-adapted strain in wild-type genome

XX Disclosure; Fig 7A-L; 36pp; English.

XX The present sequence represents a partial hepatitis A virus (HAV)

XX protein. The specification describes a DNA construct consisting

XX of a wild-type HAV genome in which the P2 region is replaced by the

XX P2 region from a cell-culture-adapted HAV strain. The construct is

XX used to demonstrate that mutations in the P2 region of a

XX cell-culture-adapted HAV strain are sufficient for establishment of

XX infection and accelerated growth in cell culture.

SO Sequence 1077 AA;

Query Match 100.0%; Score 103; DB 20; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAQEELSNEVLPPPRK 20  
DB 1033 grlkyageelsnevlppprk 1052

RESULT 10

AAK32426  
ID AAK32426 standard; Protein; 1091 AA.

XX AAK32426;

DT 17-DEC-2001 (updated)

DT 10-JUN-1993 (first entry)

XX Translated from 5' region of Hepatitis A Virus genomic clone.

XX HAV HM-175; chronic liver disease; picornavirus.

XX Hepatitis A Virus.

XX Key Location/Qualifiers

XX Region 238..1091

XX /Label= ORF

XX /note= "second putative initiation codon at

XX position 240"

XX Region 1..711

XX /note= "X's correspond to nonsense codons,

XX i.e. this region is not an ORF"

XX USN7788262-N.

XX 15-DEC-1992.

XX 30-SEP-1983; 83US-0536911.

XX 27-SEP-1984; 84US-0654942.

XX 06-OCT-1988; 88US-0256135.

XX 30-SEP-1983; 83US-0536911.

XX 06-NOV-1991; 91US-0788262.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.

XX Baltimore D, Feinstone SM;

XX Purcell RH, Racanello VR, Ticehurst JR;

XX WPI: 1993-067429/08.

XX N-PSDB; AAK36934.

XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.

XX of antigen and antibodies

XX Disclosure; Fig 7; 65pp; English.

XX HAV virion RNA was extracted from the livers of marmosets which had

XX been inoculated with HAV (the HAV had previously been passaged twice

XX in marmosets). The RNA was used to prepare ds cDNA clones by

XX standard methods. Clones contg. inserts which hybridised to RNA from

XX HAV-infected African Green Monkey Kidney cells were selected for

XX further analysis. A 7.4kb restriction map (about 99% of the HAV

XX genome) was constructed from 5 overlapping inserts. The sequence of

XX the first 3.3kb (approx.) from the 5'-terminus was determined. An

XX amino acid sequence was deduced from the entire clone and an open

XX reading frame was identified starting at position 238. A comparison

XX of the predicted HAV amino acid sequences with the known capsid

XX protein sequences of other picornaviruses (poliovirus, foot and

XX mouth disease virus and encephalomyelitis virus) revealed areas of

XX local homology.

XX (Note: Revised entry submitted to correct the patent number format of

XX US Government-owned NTIS applications to prevent clashes with ongoing US

XX granted patent numbers. For further information please visit the Derwent

XX web site at [www.derwent.com/dwpl/updates/ntis\\_us.html](http://www.derwent.com/dwpl/updates/ntis_us.html).)

SO Sequence 1091 AA;

Query Match 100.0%; Score 103; DB 14; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Hepatitis A virus.  
 XX  
 PN W08501517-A.  
 XX  
 PD 11-APR-1985.  
 XX  
 XX 27-SEP-1984; 84WO-US01552.  
 XX  
 PR 30-SEP-1983; 83US-0537911.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECH.  
 XX  
 PI Titechurst JR, Baltimore D, Feinstein SM, Purcell RH;  
 PI Racanelli VR;  
 XX  
 DR WPI: 1985-098846/16.  
 DR N-PSDB; AAN50330.  
 XX  
 PT New hepatitis A virus cDNA - useful in assays for the virus and  
 PT for prodn. of the viral antigen and antibodies to it  
 XX  
 PS Example; Fig 7; 60pp; English.  
 PS  
 XX The inventors claim HAV cDNA and a method for producing it, whereby  
 CC large amts. can be obtd. economically. The cDNA is useful in the  
 CC assay for detection of HAV quickly and easily and with high  
 CC sensitivity and specificity. The HAV cDNA is also used in the prodn.  
 CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.  
 CC  
 XX  
 SQ Sequence 854 AA;

Query Match 100.0%; Score 103; DB 6; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPpRK 20  
 |||||  
 Db 810 qrlkyageelsnevlprrk 829

RESULT 7  
 AAP50116  
 ID AAP50116 standard; Protein: 993 AA.  
 XX  
 AC AAP50116;  
 XX  
 DT 30-SEP-1991 (first entry)  
 XX  
 DE Sequence of Hepatitis A virus (HAV) immunogenic peptides  
 DE VP-1, VP-2, VP-3 and VP-4.  
 XX  
 KW Antigenic protein; immunogen; vaccine.  
 KW  
 XX Hepatitis A virus (strain CR326).  
 OS  
 XX  
 PN EPI54587-A.  
 PN  
 PD 11-SEP-1985.  
 PD  
 PE 27-FEB-1985; 85EP-0400369.  
 PE  
 PR 02-MAR-1984; 84US-0585818.  
 PR  
 PA (MERI ) MERCK & CO INC.  
 PA  
 PI Linemeyer DL, Menke JG, Reuben RG, Mitra SW;  
 PI  
 XX WPI: 1985-224964/37.  
 DR N-PSDB; AAN50139.  
 DR  
 PT New nucleotide sequences coding for hepatitis A virus antigens -  
 PT useful for eliciting normal immune response and in vaccines for

PT protecting against the virus  
 XX  
 PS Example; Page 11-17; 32pp; English.  
 XX  
 CC Within the sequence in AAN50139 is encoded the information necessary  
 CC to make the antigenic proteins of HAV. The sequences encoding for  
 CC the structural proteins begin at base 403. The key sub-unit  
 CC sequences within VP-1, designated sequences I, II, III, IV, and V,  
 CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other  
 CC nucleotide sequences which are valuable as encoding antigenic  
 CC proteins are the sequences from base 1749 to base 2722; from base  
 CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from  
 CC base 1749 to base 2722 is esp. valuable as a vector for producing  
 CC antigen protein. Sequences II-V are claimed. X in AAP50116 denotes the  
 CC translation of a stop codon.  
 CC  
 XX  
 SQ Sequence 993 AA;

Query Match 100.0%; Score 103; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPpRK 20  
 |||||  
 Db 946 qrlkyageelsnevlprrk 965

RESULT 8  
 AAP50231  
 ID AAP50231 standard; Protein: 993 AA.  
 XX  
 AC AAP50231;  
 XX  
 DT 28-NOV-1991 (first entry)  
 XX  
 DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
 DE including surface protein (VP-1).  
 XX  
 KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KW diagnostic assay.  
 KW  
 XX Hepatitis A virus.  
 OS  
 XX  
 PN Key  
 PN Location/Qualifiers  
 FH 628..993  
 FT Protein  
 FT /note- "claimed: X denotes translated stop codons  
 FT and unspecified triplets"  
 XX  
 XX EPI38704-A.  
 XX  
 PD 24-APR-1985.  
 PD  
 PE 09-OCT-1984; 84EP-0402025.  
 PE  
 PR 02-MAR-1984; 84US-0585942.  
 PR  
 PR 14-OCT-1983; 83US-0541836.  
 PR  
 PA (MERI ) MERCK & CO INC.  
 PA  
 XX Hughes JV, Scolnick EM, Tomassini JE;  
 XX  
 PI WPI: 1985-100818/17.  
 DR N-PSDB; AAN50274.  
 DR  
 XX  
 PT New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus  
 XX  
 PS Disclosure; Page 17-23; 49pp; English.  
 PS  
 XX VP1 is isolated by solubilisation of the intact virus in an ag.  
 CC antionic surfactant and a reducing agent. The viral proteins are sepd.  
 CC and the protein of molecular wt. 33000 daltons is sepd.



XX	
PF	18-APR-1997; 97WO-US06891.
XX	
PR	19-APR-1996; 96US-0015644.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI	Felds HA, Khudyakov YE;
XX	
DR	WPI; 1997-535831/49.
PT	
PT	Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an
PT	immune response to HAV in a mammal or to detect the presence of
PT	antibodies against HAV in a mammal
PS	Claim 18; Page 112; 140pp; English.
XX	
CC	Peptides AA42922-30 are immunogenic peptides corresponding to
CC	immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are
CC	substantially similar to a portion of the amino acid sequence of the P2A
CC	protein of HAV corresponding to amino acids 792-980. The present peptide
CC	is derived from amino acids 810-829, and has a reactivity of .83.38 with
CC	acute sera. Compositions containing the peptides can be used to induce an
CC	immune response to HAV in a mammal. The peptides can also be used to
CC	detect the presence of antibodies against HAV in mammalian serum. The
CC	peptides can be used to make an antibody against HAV by
CC	administering the peptide to a mammal.
XX	
Sequence	20 AA;
50	

Query Match	100.0%	Score 103	DB 18	length 20
Best Local Similarity	100.0%	Pred. No. 4.2e-09		
Matches 20	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	1	ORLKAQGEELSNVLP	PPRK	20
Db	1	qrlkyagdeelsnevlp	pprk	20

```

RESULT      2
AAB69441
ID      AAB69441 standard; Peptide; 21 AA.
AC      AAB69441;
DT      20-APR-2001 (first entry)
DE      Synthetic HAV P2A peptide, SEQ ID NO: 41.
XX
XX      Hepatitis A virus; HAV; Immunogen; Immunostimulant; virucide; vaccine.
KW      antigen; major structural capsid polypeptide; HAV antibody detection.
XX
XX      Hepatitis A virus.
OS      Synthetic.
XX
XX      WO200105824-A2.
XX
XX      25-JAN-2001.
XX
XX      14-JUL-2000; 2000MO-US19267.
XX
XX      15-JUL-1999; 99US-0144412.
XX
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX      Fields HA, Khudiyakov YE;
XX
XX      WPI; 2001-112681/12.
XX
XX      Synthetic peptides used as antigen sources for enzyme immunoassays
PT      detect anti-hepatitis A virus and as vaccines -
XX

```

PS Claim 13, Page 94; 130pp; English.

Query Match	100.0%	Score 103;	DB 22;	length 21;
Best Local Similarity	100.0%	Pred. No. 4,4e-09;		
Matches	20;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	1	ORLKTAGEELSNVLP	PPRK	20
db	1	grikYageelsnevlp	pprk	20

RESULT	3
AAW42930	
ID	AAW42930 standard; peptide; 25 AA.
XX	
AC	AAW42930;
XX	
DT	28-APR-1998 (first entry)
XX	
DE	Immunogenic Hepatitis A virus peptide YK-1665.
XX	
KW	Immunogenic peptide; immunogenic epitope; P2A protein;
XX	Immune response; antibody.
XX	
OS	Synthetic.
XX	Hepatitis A virus.
EN	MO9740147-A1.
XX	
PD	30-OCT-1997.
XX	
PF	18-APR-1997; 97MO-US06891.
XX	
PR	19-APR-1996; 96US-0005644.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Fields HA, Khudjakov YE;
XX	
DR	WPI; 1997-535831/49.
XX	
PT	Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an
PT	immune response to HAV in a mammal or to detect the presence of
PT	antibodies against HAV in a mammal
XX	
PS	Claim 18; Page 112; 140pp; English.
XX	
CC	Peptides AAW42922-30 are immunogenic peptides corresponding to
CC	immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are
CC	substantially similar to a portion of the amino acid sequence of the P2A
CC	protein of HAV corresponding to amino acids 792-980. Compositions
CC	containing the peptides can be used to induce an immune response to HAV
CC	in a mammal. The peptides can also be used to detect the presence of

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:01:42 ; Search time 209.1 Seconds  
(without alignments)  
10.624 Million cell updates/sec

Title: US-09-171-432a-41  
Perfect score: 103  
Sequence: 1 QRLKYAOEELSNEYLPPEPRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

A.Geneseq\_032802:\*

1: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT:\*  
2: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:\*  
3: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:\*  
4: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:\*  
5: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:\*  
6: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:\*  
7: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:\*  
8: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:\*  
9: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:\*  
10: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:\*  
11: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:\*  
12: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:\*  
13: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:\*  
14: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:\*  
15: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:\*  
16: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:\*  
17: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:\*  
18: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:\*  
19: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:\*  
20: /SID55/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:\*  
21: /SID55/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:\*  
22: /SID55/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	20	18 AAM42924	Immunogenic Hepati
2	103	100.0	21	22 AAB69441	Synthetic HAV P2A
3	103	100.0	25	18 AAM42930	Immunogenic Hepati
4	103	100.0	26	22 AAB69447	Synthetic HAV P2A
5	103	100.0	366	6 AAP50230	Sequence of hepati
6	103	100.0	854	6 AAP50287	Sequence encoded b
7	103	100.0	993	6 AAP50116	Sequence of Hepati
8	103	100.0	993	6 AAP50231	Sequence encoded b
9	103	100.0	1077	20 AAM95559	A partial hepatitis
10	103	100.0	1091	14 AAR32426	Translated from 5'
11	103	100.0	2227	7 AAP60066	Sequence of viral

12	103	100.0	2227	11 AAR05697	Attenuated hepatitis
13	103	100.0	2227	18 AAM34074	Hepatitis A virus
14	103	100.0	2227	21 AAB18607	Amino acid sequenc
15	103	100.0	2227	21 AAB18609	Amino acid sequenc
16	98	95.1	2227	21 AAB18608	Amino acid sequenc
17	94	91.3	839	12 AAR15629	Capsid region of c
18	50	48.5	150	22 ABB68507	Drosophila melanog
19	48	46.6	444	18 AAM20624	H. pylori cytoplas
20	48	45.6	1176	22 ABB50592	Drosophila melanog
21	47	45.6	221	22 ABB68504	Drosophila melanog
22	45	43.7	20	18 AAM42923	Immunogenic Hepati
23	45	43.7	21	22 AAB69440	Synthetic HAV P2A
24	45	43.7	263	22 ABB31812	Peptide #4463 enco
25	45	43.7	263	22 ABB37043	Peptide #4549 enco
26	45	43.7	263	22 ABB22358	Protein #4357 enco
27	45	43.7	263	22 AAM57770	Human brain expres
28	45	43.7	263	22 AAM70185	Human bone marrow
29	45	43.7	263	22 AAM18008	Peptide #4442 enco
30	45	43.7	263	22 AAM30519	Peptide #4556 enco
31	45	43.7	305	22 AAM05648	Peptide #4330 enco
32	44	42.7	334	22 AAM41677	Human polypeptide
33	44	42.7	334	22 ABB26758	Novel human diagno
34	44	42.7	334	22 ABB26759	Novel human diagno
35	44	42.7	368	22 ABB28311	Novel human diagno
36	44	42.7	420	22 ABB63362	Drosophila melanog
37	44	42.7	450	22 ABB26763	Novel human diagno
38	44	42.7	633	22 ABB28306	Novel human diagno
39	44	42.7	755	22 ABB59259	Drosophila melanog
40	44	42.7	914	16 AAR77274	ORC1 subunit of ye
41	44	42.7	914	18 AAM22224	S. cerevisiae orig
42	43.5	42.2	1091	10 AAP98500	Partial sequence e
43	43.5	42.2	3685	10 AAP90290	Human Duchenne mus
44	43.5	42.2	3685	10 AAP90373	Sequence encoded b
45	43.5	41.7	205	22 AAE07109	Human gene 6 enco
46	43	41.7	213	22 AAE07083	Human gene 6 enco
47	43	41.7	302	21 AAG06817	Arabidopsis thalia
48	43	41.7	362	21 AAG06816	Arabidopsis thalia
49	43	41.7	372	21 AAG06815	Arabidopsis thalia
50	43	41.7	390	20 AAV05302	S. aureus protein
51	43	41.7	585	18 AAM01671	Influenza B/Panama
52	43	41.7	585	20 AAM75443	Influenza virus B/
53	43	41.7	585	22 AAE04953	Influenza virus B/
54	43	41.7	586	18 AAM01675	Influenza B/Harbin
55	43	41.7	586	20 AAM75447	Influenza virus B/
56	43	41.7	586	22 AAE04957	Influenza virus B/
57	43	41.7	589	18 AAM01672	Influenza B/Nethe
58	43	41.7	589	22 AAM75444	Influenza virus B/
59	43	41.7	589	22 AAE04954	Influenza virus B/
60	43	41.7	592	18 AAM01674	Influenza A/Shanha

#### ALIGNMENTS

RESULT	1
ID	AAM42924
AAW42924	standard; peptide: 20 AA.
XX	
AC	AAW42924;
XX	
DT	28-APR-1998 (first entry)
XX	
DE	Immunogenic Hepatitis A virus peptide YK-1317.
XX	
KW	Immunogenic peptide; immunogenic epitope; P2A protein;
KM	Immune response; antibody.
XX	
OS	Synthetic.
OS	Hepatitis A virus.
XX	
PN	WO9740147-A1.
XX	
PD	30-OCT-1997.

Attenuated hepatitis  
Hepatitis A virus  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Capsid region of c  
Drosophila melanog  
H. pylori cytoplas  
Drosophila melanog  
Drosophila melanog  
Immunogenic Hepati  
Synthetic HAV P2A  
Peptide #4463 enco  
Peptide #4549 enco  
Protein #4357 enco  
Human brain expres  
Human bone marrow  
Peptide #4442 enco  
Peptide #4556 enco  
Peptide #4330 enco  
Human polypeptide  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Drosophila melanog  
ORC1 subunit of ye  
S. cerevisiae orig  
Partial sequence e  
Human Duchenne mus  
Sequence encoded b  
Human gene 6 enco  
Human gene 6 enco  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
S. aureus protein  
Influenza B/Panama  
Influenza virus B/  
Influenza virus B/  
Influenza B/Harbin  
Influenza virus B/  
Influenza virus B/  
Influenza B/Nethe  
Influenza virus B/  
Influenza virus B/  
Influenza A/Shanha

DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 NCBI\_TaxID=12092;  
 RX 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A16;  
 RA Fujiwara K.;  
 RT "Hepatitis A virus VP1/2A junction."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB046881; BAB08038.1; -  
 DR InterPro; IPR000886; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT 56  
 SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AFA3C CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. NO. 3.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLVVGKQRLKYAOEE 20  
 |||  
 DB 36 PYKELRLVVGKQRLKYAOEE 55

Search completed: June 16, 2002, 00:08:51  
 Job time: 789 sec

DT 01-DEC-2001 (TREMBLER, 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISR-70, AKKO DISTRICT, ISRAEL, 1993;  
RA Karethy Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77247; CAB01040.1; -;  
DR InterPro; IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGRKRLKYAEE 20  
Db 36 PYKELRLEVGRKRLKYAEE 55

RESULT 22  
ID 067823 PRELIMINARY; PRT; 56 AA.  
AC 067823;  
DT 01-NOV-1996 (TREMBLER, 01, Created)  
DT 01-NOV-1996 (TREMBLER, 01, last sequence update)  
DT 01-DEC-2001 (TREMBLER, 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISR-79, AKKO DISTRICT, ISRAEL, 1993;  
RA Karethy Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77248; CAB01041.1; -;  
DR InterPro; IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGRKRLKYAEE 20  
Db 36 PYKELRLEVGRKRLKYAEE 55

RESULT 23

Q91PD2  
ID Q91PD2 PRELIMINARY; PRT; 56 AA.  
AC Q91PD2;  
DT 01-OCT-2000 (TREMBLER, 15, Created)  
DT 01-OCT-2000 (TREMBLER, 15, last sequence update)  
DT 01-JUN-2001 (TREMBLER, 17, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A10;  
RA Fujiwara K.;  
RT "Hepatitis A virus VP1/2A junction."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB046878; BAB08035.1; -;  
DR InterPro; IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGRKRLKYAEE 20  
Db 36 PYKELRLEVGRKRLKYAEE 55

RESULT 24  
ID Q91PD1 PRELIMINARY; PRT; 56 AA.  
AC Q91PD1;  
DT 01-OCT-2000 (TREMBLER, 15, Created)  
DT 01-OCT-2000 (TREMBLER, 15, last sequence update)  
DT 01-JUN-2001 (TREMBLER, 17, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A157;  
RA Fujiwara K.;  
RT "Hepatitis A virus VP1/2A junction."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB046879; BAB08036.1; -;  
DR InterPro; IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGRKRLKYAEE 20  
Db 36 PYKELRLEVGRKRLKYAEE 55

RESULT 25  
Q91PC9 PRELIMINARY; PRT; 56 AA.  
ID Q91PC9  
AC Q91PC9;



RA Mendelson E.:  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 277243; CAB01036.1; -.  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEARLEVGKORLKYAOE 20  
Db 36 PYKEARLEVGKORLKYAOE 55

RESULT 18  
O67819 PRELIMINARY; PRT; 56 AA.  
AC 067819;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-10, AFULA DISTRICT, ISRAEL, 1993);  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.:  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 277244; CAB01037.1; -.  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEARLEVGKORLKYAOE 20  
Db 36 PYKEARLEVGKORLKYAOE 55

RESULT 19  
O67820 PRELIMINARY; PRT; 56 AA.  
AC 067820;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.

OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-12, AFULA DISTRICT, ISRAEL, 1993);  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.:  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 277245; CAB01038.1; -.  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEARLEVGKORLKYAOE 20  
Db 36 PYKEARLEVGKORLKYAOE 55

RESULT 20  
O67821 PRELIMINARY; PRT; 56 AA.  
AC 067821;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-15, AFULA DISTRICT, ISRAEL, 1993);  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.:  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 277246; CAB01039.1; -.  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEARLEVGKORLKYAOE 20  
Db 36 PYKEARLEVGKORLKYAOE 55

RESULT 21  
O67822 PRELIMINARY; PRT; 56 AA.  
AC 067822;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

```
RT 1982-1996.";
RL J. Med. Virol. 51:273-279(1997).
DR EMBL: U68697; AAB53593.1; -.
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT SEQUENCE 56 AA; 6642 MW; 465CP51846AEF4BC CRC64;
SQ

Query Match
Best Local Similarity 100.0%; Score 102; DB 12; Length 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLEVGRKRLKYAOEE 20
Db 36 PYKELRLEVGRKRLKYAOEE 55

RESULT 14
O39873 PRELIMINARY; PRT; 56 AA.
AC O39873;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2333;
RX MEDLINE=97247817; PubMed=9093940;
RA Taylor M.B.;
RT "Molecular epidemiology of South African strains of hepatitis A virus:
RL J. Med. Virol. 51:273-279(1997).
DR EMBL: U68698; AAB53594.1; -.
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;
SQ

Query Match
Best Local Similarity 100.0%; Score 102; DB 12; Length 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLEVGRKRLKYAOEE 20
Db 36 PYKELRLEVGRKRLKYAOEE 55

RESULT 15
O39874 PRELIMINARY; PRT; 56 AA.
AC O39874;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=314274;
RX MEDLINE=97247817; PubMed=9093940;
RA Taylor M.B.;
```

```
RT "Molecular epidemiology of South African strains of hepatitis A virus:
RL J. Med. Virol. 51:273-279(1997).
DR EMBL: U68699; AAB53595.1; -.
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;
SQ

Query Match
Best Local Similarity 100.0%; Score 102; DB 12; Length 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLEVGRKRLKYAOEE 20
Db 36 PYKELRLEVGRKRLKYAOEE 55

RESULT 16
O96830 PRELIMINARY; PRT; 56 AA.
AC O96830;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=504184;
RX MEDLINE=97247817; PubMed=9093940;
RA Taylor M.B.;
RT "Molecular epidemiology of South African strains of hepatitis A virus:
RL J. Med. Virol. 51:273-279(1997).
DR EMBL: U68700; AAB53596.1; -.
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT SEQUENCE 56 AA; 6598 MW; 7528C51846AEF4B9 CRC64;
SQ

Query Match
Best Local Similarity 100.0%; Score 102; DB 12; Length 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLEVGRKRLKYAOEE 20
Db 36 PYKELRLEVGRKRLKYAOEE 55

RESULT 17
O67818 PRELIMINARY; PRT; 56 AA.
AC O67818;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISR-05; AFOLA DISTRICT, ISRAEL, 1993;
RA Karetnyi Y.V., Shulman L.M., Manor J., Lettner L., Shehab S.,
```

DR EMBL: U68693; AAB53589.1; -  
 DR InterPro: IPR000886; ER\_TARGET.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA: 6614 MW: 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLKYAOE 20  
 |||||  
 DB 36 PYKEIRLEVGKORLKYAOE 55

RESULT 10  
 ID 039869 PRELIMINARY; PRT; 56 AA.  
 AC 039869;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9600382;  
 RX MEDLINE-97247817; PubMed-9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL: U68694; AAB53590.1; -  
 DR InterPro: IPR000886; ER\_TARGET.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA: 6614 MW: 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLKYAOE 20  
 |||||  
 DB 36 PYKEIRLEVGKORLKYAOE 55

RESULT 11  
 ID 039870 PRELIMINARY; PRT; 56 AA.  
 AC 039870;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-406808;  
 RX MEDLINE-97247817; PubMed-9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";

RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL: U68695; AAB53591.1; -  
 DR InterPro: IPR000886; ER\_TARGET.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA: 6642 MW: 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLKYAOE 20  
 |||||  
 DB 36 PYKEIRLEVGKORLKYAOE 55

RESULT 12  
 ID 039871 PRELIMINARY; PRT; 56 AA.  
 AC 039871;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-923359;  
 RX MEDLINE-97247817; PubMed-9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL: U68696; AAB53592.1; -  
 DR InterPro: IPR000886; ER\_TARGET.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA: 6642 MW: 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLKYAOE 20  
 |||||  
 DB 36 PYKEIRLEVGKORLKYAOE 55

RESULT 13  
 ID 039872 PRELIMINARY; PRT; 56 AA.  
 AC 039872;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VDM;  
 RX MEDLINE-97247817; PubMed-9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:"

DR PROSITE: PS00014; ER\_TARGET: UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVKGKRLKYAOEE 20  
DB 36 PYKELRLEVKGKRLKYAOEE 55

## RESULT 6

039866 PRELIMINARY; PRT; 56 AA.  
AC 039866; 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=96001190;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL: U68690; AAB53586.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET: UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVKGKRLKYAOEE 20  
DB 36 PYKELRLEVKGKRLKYAOEE 55

## RESULT 7

039829 PRELIMINARY; PRT; 56 AA.  
AC 039829; 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=406909;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL: U68691; AAB53587.1; -;

DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET: UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6623 MW; A917151846AEFA47 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVKGKRLKYAOEE 20  
DB 36 PYKELRLEVKGKRLKYAOEE 55

## RESULT 8

039867 PRELIMINARY; PRT; 56 AA.  
AC 039867; 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=JVR;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL: U68692; AAB53588.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET: UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVKGKRLKYAOEE 20  
DB 36 PYKELRLEVKGKRLKYAOEE 55

## RESULT 9

039868 PRELIMINARY; PRT; 56 AA.  
AC 039868; 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=503712;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).

FT NON\_TER 55 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

## Query Match

Best Local Similarity 100.0%; Score 102; DB 12; Length 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKVAOE 20  
Db 35 PYKELRLEVGKORLKVAOE 54

## RESULT 2

0999T4 ID 0999T4 PRELIMINARY; PRT; 55 AA.

AC 0999T4;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHILE-3;  
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Collina R.,  
RA Billand S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ06384; CAC2923.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
SQ

## Query Match

Best Local Similarity 100.0%; Score 102; DB 12; Length 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKVAOE 20  
Db 35 PYKELRLEVGKORLKVAOE 54

## RESULT 3

0999T3 ID 0999T3 PRELIMINARY; PRT; 55 AA.

AC 0999T3;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHILE-9;  
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Collina R.,  
RA Billand S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ06385; CAC2923.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.

FT NON\_TER 1 1  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match  
Best Local Similarity 100.0%; Score 102; DB 12; Length 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKVAOE 20  
Db 35 PYKELRLEVGKORLKVAOE 54

## RESULT 4

0999T2 ID 0999T2 PRELIMINARY; PRT; 55 AA.

AC 0999T2;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URU3;  
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Collina R.,  
RA Billand S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ06386; CAC2923.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
SQ

Query Match  
Best Local Similarity 100.0%; Score 102; DB 12; Length 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKVAOE 20  
Db 35 PYKELRLEVGKORLKVAOE 54

## RESULT 5

039865 ID 039865 PRELIMINARY; PRT; 56 AA.

AC 039865;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=412991;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL: U06869; AAB53585.1; -;  
DR InterPro: IPR000886; ER\_target.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:51 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432A-40

Perfect score: 102

Sequence: 1 PYKELRLVKGQRKRYAQEE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

1: SP:archaea:\*  
2: SP:bacteria:\*  
3: SP:fungi:\*  
4: SP:human:\*  
5: SP:invertebrate:\*  
6: SP:mammal:\*  
7: SP:mnc:\*  
8: SP:organelle:\*  
9: SP:phage:\*  
10: SP:plant:\*  
11: SP:rodent:\*  
12: SP:virus:\*  
13: SP:vertebrate:\*  
14: SP:unclassified:\*  
15: SP:tvirus:\*  
16: SP:bacteriophage:\*  
17: SP:archaeoprotein:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102	100.0	55	12	Q99908 hepatitis a
2	102	100.0	55	12	Q99904 hepatitis a
3	102	100.0	55	12	Q99903 hepatitis a
4	102	100.0	55	12	Q99902 hepatitis a
5	102	100.0	56	12	Q99865 hepatitis a
6	102	100.0	56	12	Q99866 hepatitis a
7	102	100.0	56	12	Q99829 hepatitis a
8	102	100.0	56	12	Q99867 hepatitis a
9	102	100.0	56	12	Q99868 hepatitis a
10	102	100.0	56	12	Q99869 hepatitis a
11	102	100.0	56	12	Q99870 hepatitis a
12	102	100.0	56	12	Q99871 hepatitis a
13	102	100.0	56	12	Q99872 hepatitis a
14	102	100.0	56	12	Q99873 hepatitis a
15	102	100.0	56	12	Q99874 hepatitis a
16	102	100.0	56	12	Q99830 hepatitis a

## ALIGNMENTS

17	102	100.0	56	12	Q67818 hepatitis a
18	102	100.0	56	12	Q67819 hepatitis a
19	102	100.0	56	12	Q67820 hepatitis a
20	102	100.0	56	12	Q67821 hepatitis a
21	102	100.0	56	12	Q67822 hepatitis a
22	102	100.0	56	12	Q67823 hepatitis a
23	102	100.0	56	12	Q91P02 hepatitis a
24	102	100.0	56	12	Q91P01 hepatitis a
25	102	100.0	56	12	Q91P09 hepatitis a
26	102	100.0	56	12	Q91P08 hepatitis a
27	102	100.0	56	12	Q91P07 hepatitis a
28	102	100.0	56	12	Q91P05 hepatitis a
29	102	100.0	56	12	Q91P04 hepatitis a
30	102	100.0	56	12	Q91P03 hepatitis a
31	102	100.0	56	12	Q91P02 hepatitis a
32	102	100.0	56	12	Q91P01 hepatitis a
33	102	100.0	56	12	Q91P00 hepatitis a
34	102	100.0	56	12	Q91P99 hepatitis a
35	102	100.0	56	12	Q91P98 hepatitis a
36	102	100.0	56	12	Q91P97 hepatitis a
37	102	100.0	56	12	Q91P96 hepatitis a
38	102	100.0	56	12	Q91P95 hepatitis a
39	102	100.0	56	12	Q91P94 hepatitis a
40	102	100.0	56	12	Q91P93 hepatitis a
41	102	100.0	56	12	Q91P92 hepatitis a
42	102	100.0	56	12	Q91P91 hepatitis a
43	102	100.0	56	12	Q91P90 hepatitis a
44	102	100.0	56	12	Q91P89 hepatitis a
45	102	100.0	56	12	Q91P88 hepatitis a
46	102	100.0	56	12	Q91P87 hepatitis a
47	102	100.0	56	12	Q91P86 hepatitis a
48	102	100.0	56	12	Q91P85 hepatitis a
49	102	100.0	56	12	Q91P84 hepatitis a
50	102	100.0	56	12	Q91P83 hepatitis a
51	102	100.0	56	12	Q91P82 hepatitis a
52	102	100.0	56	12	Q91P81 hepatitis a
53	102	100.0	56	12	Q91P80 hepatitis a
54	102	100.0	56	12	Q91P79 hepatitis a
55	102	100.0	56	12	Q91P78 hepatitis a
56	102	100.0	56	12	Q91P77 hepatitis a
57	102	100.0	56	12	Q91P76 hepatitis a
58	102	100.0	56	12	Q91P75 hepatitis a
59	102	100.0	56	12	Q91P74 hepatitis a
60	102	100.0	56	12	Q91P73 hepatitis a

RESULT 1  
ID Q99908 PRELIMINARY; PRT; 55 AA.  
AC Q99908;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-ARG-6:  
RA Costa-Matelloi M., Ferre V., Monphoeh S., Garcia L., Colina R.,  
Billaud S., Vega T., Perez-Bercoff R., Cristina J.,  
"Genetic Variability of Hepatitis A Virus in South America reveals  
RT heterogeneity and co-circulation during epidemic outbreaks".  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306370; CAC29219.1; -;  
DR InterPro; IPR000886; ER-target.  
DR PROSITE; PS00014; ER-TARGET; UNKNOWN\_1.  
FT NON\_TER 1



FT LIPID 65 65 GERANYL-GERANYL.  
FT PROPEP 66 68 REMOVED IN MATURE FORM (BY SIMILARITY).  
SQ SEQUENCE 68 AA; 7205 MW; E3EC3CEA0EB0BD5A CRC64;

Query Match 40.2%; Score 41; DB 1; Length 68;  
Best Local Similarity 43.8%; Pred. No. 5.5;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 KEKLEVGKORLKYAQ 18  
Db 15 EQLKLENGVERIKYSQ 30

RESULT 24  
GST2\_HUMAN STANDARD; PRT; 147 AA.  
ID GST2\_HUMAN  
AC 099735;  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Microsomal glutathione S-transferase 2 (EC 2.5.1.18) (Microsomal GST-2) (Microsomal GST-II).  
GN MGST2 OR GST2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96355624; PubMed=8703034;  
RA Jacobson P.J., Mancini J.A., Ford-Hutchinson A.W.;  
RT "Identification and characterization of a novel human microsomal glutathione S-transferase with leukotriene C4 synthase activity and significant sequence identity to 5-lipoxygenase-activating protein and leukotriene C4 synthase";  
RT J. Biol. Chem. 271:22203-22210(1996).  
CC -1- FUNCTION: CAN CATALYZE THE PRODUCTION OF LTC4 FROM LTC4 AND REDUCED GLUTATHIONE. CAN CATALYZE THE CONJUGATION OF 1-CHLORO-2,4-DINITROBENZENE WITH REDUCED GLUTATHIONE.  
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.  
CC -1- SUBUNIT: HOMOTRIMER.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- TISSUE SPECIFICITY: LIVER, SPLEEN, SKELETAL MUSCLE, HEART, ADRENALS, PANCREAS, PROSTATE, TESTIS, FETAL LIVER, AND FETAL SPLEEN. VERY LOW EXPRESSION IN LUNG, BRAIN, PLACENTA, AND BONE MARROW.  
CC -1- SIMILARITY: BELONGS TO THE MAPEG FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; U77604; AAC51768.1; -  
DR MIM; 601733; -  
DR InterPro; IPR001129; MAPEG.  
DR Pfam; PF011124; MAPEG; 1.  
KW PROSITE; PS01297; FLAP\_GST2\_LTC4S; 1.  
KW Transferase; Transmembrane; Microsome; Leukotriene biosynthesis.  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 59 79 POTENTIAL.  
FT TRANSMEM 111 131 POTENTIAL.  
SQ SEQUENCE 147 AA; 16620 MW; D0E89B46885D16EF CRC64;

Query Match 40.2%; Score 41; DB 1; Length 147;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 LEVGKORLKY 16  
Db 24 LQVGKARLKY 33

RESULT 25  
Y234\_AQUAE STANDARD; PRT; 318 AA.  
ID Y234\_AQUAE  
AC 066423;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AA34.  
GN AA34.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";  
RT Nature 392:353-358(1998).  
RL Nature 392:353-358(1998).  
CC -1- SIMILARITY: STRONG. TO A AEOLICUS AN07 AND AA11.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AE000667; AAC07975.1; -  
KW Hypothetical protein; Plasmid; Complete proteome.  
SQ SEQUENCE 318 AA; 37360 MW; EDB125C719DE49A3 CRC64;

Query Match 40.2%; Score 41; DB 1; Length 318;  
Best Local Similarity 47.4%; Pred. No. 28;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 YKELRLEVGKORLKYAE 20  
Db 55 FEETLALMKORLKYEE 73

Search completed: June 16, 2002, 00:10:06  
Job time: 654 sec



RA Wang Y.-X., Chen S.-J., Chen Z.;  
 RT "Identification of genes expressed in human CD34(+) hematopoietic  
 RT stem/progenitor cells by expressed sequence tags and efficient full-  
 RT length cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human; TISSUE-Brain;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Bovine, and Rat; TISSUE-Liver;  
 RX MEDLINE-92195304; PubMed-1549114;  
 RA Fisher K.J., Aronson N.N. Jr.;  
 RT "Characterization of the cDNA and genomic sequence of a G protein  
 RT gamma subunit (gamma 5)";  
 RL Mol. Cell. Biol. 12:1585-1591(1992).  
 RN [5]  
 RP SEQUENCE.  
 RC SPECIES-Bovine; TISSUE-Spleen;  
 RX MEDLINE-9356792; PubMed-8352779;  
 RA Morishita R., Masuda K., Niwa M., Kato K., Asano T.;  
 RT "Identification of three forms of the gamma subunit of G proteins  
 RT isolated from bovine spleen.";  
 RL Biochem. Biophys. Res. Commun. 194:1221-1227(1993).  
 RN [6]  
 RP SEQUENCE OF 8-53 FROM N.A.  
 RC SPECIES-Mouse; STRAIN-CE-1 / HARLAN;  
 RX MEDLINE-97011591; PubMed-8858601;  
 RA Williams C.J., Schultz R.M., Kopf G.S.;  
 RT "G protein gene expression during mouse oocyte growth and maturation,  
 RT and preimplantation embryo development.";  
 RL Mol. Reprod. Dev. 44:315-323(1996).  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 CC EFFECTOR INTERACTION.  
 CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST SEVEN DIFFERENT GAMMA SUBUNITS  
 CC WITH POSSIBLE TISSUE-SPECIFIC DISTRIBUTION.  
 CC -1- SIMILARITY: BELONGS TO THE G PROTEIN GAMMA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF085709; AAC72203.1; -  
 DR EMBL; AF085708; AAC72203.1; JOINED.  
 DR EMBL; AF085708; AAC72203.1; -  
 DR EMBL; AF085708; AAC72203.1; -  
 DR EMBL; BC003563; AAH03563.1; -  
 DR EMBL; M95779; AAA30535.1; -  
 DR EMBL; M95780; AAA41188.1; -  
 DR EMBL; U38498; AAB01729.1; -  
 DR PIR; JN0701; JN0701.  
 DR PIR; B42243; B42243.  
 DR HSSP; P16874; I6G2.  
 DR MIM; 600874; -  
 DR MGD; MGI:109164; Gng5.  
 DR InterPro; IPR001770; G-gamma.  
 DR Pfam; PF00631; G-gamma; 1.  
 DR PRINTS; PR00321; GPROTEIN.  
 DR PRODOM; PD003783; G-gamma; 1.  
 DR SMART; SM00224; GGL; 1.  
 DR PROSITE; PSS0058; G-PROTEIN\_GAMMA; 1.  
 KW Transducer; Prenylation; Lipoprotein; Multigene family.  
 FT LIPID 65 65 GERANYL-GERANYL (BY SIMILARITY).

FT PROPEP 66 68 REMOVED IN MATURE FORM (BY SIMILARITY).  
 SQ SEQUENCE 68 AA; 7318 MW; 9AF7A1655863602 CRC64;  
 Query Match 40.2%; Score 41; DB 1; Length 68;  
 Best Local Similarity 50.0%; Pred. No. 5.3;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 0Y 3 KRLREYKORLKYAQ 18  
 Db 15 QQLRLERGLNRKVSQ 30  
 RESULT 23  
 GBGA\_HUMAN STANDARD; PRT; 68 AA.  
 ID GBGA\_HUMAN  
 AC P50151;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-10 subunit.  
 GN GNG10 OR GNGR10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE-95394940; PubMed-7665596.  
 RA Ray K., Kunsch C., Bonner L.M., Robishaw J.D.;  
 RT "Isolation of cDNA clones encoding eight different human G protein  
 RT gamma subunits, including three novel forms designated the gamma 4,  
 RT gamma 10, and gamma 11 subunits.";  
 RL J. Biol. Chem. 270:21765-21771(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Laird G.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 CC EFFECTOR INTERACTION. INTERACTS WITH BETA-1 AND BETA-2, BUT NOT  
 CC WITH BETA-3.  
 CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -1- TISSUE SPECIFICITY: ABUNDANTLY AND UBICITOUSLY EXPRESSED.  
 CC -1- SIMILARITY: BELONGS TO THE G PROTEIN GAMMA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U31383; AAC50205.1; -  
 DR EMBL; AL135787; CAC22155.1; -  
 DR EMBL; BC010384; AAH10384.1; -  
 DR HSSP; P16874; I6P2.  
 DR MIM; 604389; -  
 DR InterPro; IPR001770; G-gamma.  
 DR Pfam; PF00631; G-gamma; 1.  
 DR PRINTS; PR00321; GPROTEIN.  
 DR PRODOM; PD003783; G-gamma; 1.  
 DR SMART; SM00224; GGL; 1.  
 DR PROSITE; PSS0058; G-PROTEIN\_GAMMA; 1.  
 KW Transducer; Prenylation; Lipoprotein; Multigene family.

RX MEDLINE-92089025; PubMed-1751451;  
RA Luttinger A.L., Springer A.L., Schmid M.B.;  
RT "A cluster of genes that affects nucleoid segregation in Salmonella  
typhimurium.";  
RL New Biol. 3:687-697(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;  
RA MEDLINE-21534948; PubMed-11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RL Nature 413:852-856(2001).  
CC -1- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME  
CC SEPARATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.  
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION  
CC OF A CIRCULAR DNA MOLECULE.  
CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M68936; AAA27180.1; -  
DR EMBL: AE008846; AAL22048.1; -  
DR PIR: A45582; A45582.  
DR HSSP: P09097; IAB4.  
DR StyGene: SG10276; parC.  
DR InterPro: IPR002205; DNA\_topoisomIV.  
DR Pfam: PF00521; DNA\_topoisomIV; 1.  
DR SMART: SM00434; TOP4c; 1.  
KW Topoisomerase; Isomerase; DNA-binding; Complete proteome.  
FT ACT\_SITE 120 120 DNA\_CLEAVAGE (BY SIMILARITY).  
FT CONFLICT 241 242 MR -> IG (IN REF. 1).  
SQ SEQUENCE 752 AA; 84037 MW; F34FD7FFD206760 CRC64;

Query Match 41.2%; Score 42; DB 1; Length 752;  
Best Local Similarity 40.0%; Pred. No. 49;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 PYKELFLEVGKORLKYAOEE 20  
DB 694 POSTLTIHVGKRIKIRPEE 713  
| : ||| : |||  
| : ||| : |||

RESULT 21  
SVL\_CHLTR STANDARD: PRT; 939 AA.  
AC 084304;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (VALRS).  
GN VALS OR C302.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-D/TW-3/CX;  
RA MEDLINE-99000609; PubMed-9784136;  
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) -> AMP + diphosphate  
CC + L-valyl-tRNA(Val).  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AE001302; AAC67895.1; -  
DR HSSP: P96142; IGAX.  
DR InterPro: IPR002300; tRNA-synt\_1a.  
DR InterPro: IPR001412; tRNA-synt\_1.  
DR InterPro: IPR002303; tRNA-synt\_val.  
DR Pfam: PF00133; tRNA-synt\_1; 1.  
DR PRINTS: PR00986; TRNASYNTVAL.  
DR PROSITE: PS00178; AA-TRNA\_LIGASE\_I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 47 57 "HIGH" REGION.  
FT SITE 563 567 "KMSKS" REGION.  
FT BINDING 566 566 ATP (BY SIMILARITY).  
SQ SEQUENCE 939 AA; 107036 MW; CEB8449DC7BB9066 CRC64;

Query Match 41.2%; Score 42; DB 1; Length 939;  
Best Local Similarity 38.9%; Pred. No. 62;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 KELLFLEVGKORLKYAOEE 20  
DB 97 RHLKASLKGKRTDFSRRE 114  
| : ||| : ||| : |||  
| : ||| : ||| : |||

RESULT 22  
GBGS\_HUMAN STANDARD: PRT; 68 AA.  
AC P30670; O61015; (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Guanine nucleotide-binding protein G(I)/G(s)/G(o) gamma-5 subunit.  
GN GNG5 OR GNGT5.  
OS Homo sapiens (Human).  
OS Mus musculus (Mouse).  
OS Rattus norvegicus (Rat), and  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606, 10090, 10116, 9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RA MEDLINE-99009227; PubMed-9790912;  
RA Liu B., Aronson N.N. J.;  
RT "Structure of human G protein gamma5 gene GNG5.";  
RL Biochem. Biophys. Res. Commun. 251:88-94(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human; TISSUE=Blood;  
RA MEDLINE-96318631; PubMed-9653160;  
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,  
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 179 AA; 20230 MW; ABB6FE42D5BEBE83 CRC64;  
Query Match 41.7%; Score 42.5; DB 1; Length 179;  
Best Local Similarity 61.1%; Pred. No. 8.9;  
Matches 11; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
QY 4 ELR-LEVGKORLKYAOEE 20  
||| ||| :||| :|||  
Db 104 ELRSLKIKOKLAKKE 121  
RESULT 19  
PARC\_ECOLI STANDARD; PRT; 752 AA.  
AC P20082;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Topoisomerase IV subunit A (EC 5.99.1.-).  
GN PARC OR B3019 OR Z4373 OR ECS3903.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=91004247; PubMed=2170028;  
RA Kato J.-I., Nishimura Y., Imanura R., Niki H., Hiraga S., Suzuki H.;  
RT "New topoisomerase essential for chromosome segregation in E. coli.";  
RL Cell 63:393-404(1990).  
RN [2]  
RP ERRATUM.  
RA Kato J.-I., Nishimura Y., Imanura R., Niki H., Hiraga S., Suzuki H.;  
RL Cell 65:1289-1290(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potlowski K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,  
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RN [6]  
RP SEQUENCE OF 724-752 FROM N.A.

RX MEDLINE=92212294; PubMed=1557036;  
RA Coleman J.;  
RT "Characterization of the Escherichia coli gene for  
RT 1-acyl-sn-glycerol-3-phosphate acyltransferase (plsc).";  
RL Mol. Gen. Genet. 232:295-303(1992).  
RN [7]  
RP REVIEWS, AND CHARACTERIZATION.  
RC STRAIN-K12;  
RX MEDLINE=94043292; PubMed=8227000;  
RA Peng H., Mariani K.J.;  
RT "Escherichia coli topoisomerase IV. Purification, characterization,  
RT subunit structure, and subunit interactions.";  
RL J. Biol. Chem. 268:24481-24481(1993).  
CC -1- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME  
CC SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.  
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION  
CC OF A CIRCULAR DNA MOLECULE.  
CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M58408; AAA24297.1; ALT\_INIT.  
DR EMBL; M63491; AAA24396.1; -.  
DR EMBL; U28377; AAA69187.1; -.  
DR EMBL; AE000384; AAC76055.1; -.  
DR EMBL; AE005531; AAG58155.1; -.  
DR EMBL; AP002563; BAB37326.1; -.  
DR EMBL; L22025; AAC36840.1; -.  
DR PIR; A36075; A36075.  
DR PIR; A36075; A36075.  
DR HSSP; P09097; IAB4.  
DR Ecogene; EGI0686; parC.  
DR InterPro: IPR002205; DNA\_topoisolv.  
DR Pfam; PF00521; DNA\_topoisolv.  
DR SMART; SM00434; TOP4C; 1.  
DR Topoisomerase; Isomerase; DNA-binding; Complete proteome.  
FT ACT\_SITE 120 120 DNA\_CLEAVAGE (BY SIMILARITY).  
SQ SEQUENCE 752 AA; 83831 MW; 0D4907E96CEB7086 CRC64;  
Query Match 41.2%; Score 42; DB 1; Length 752;  
Best Local Similarity 40.0%; Pred. No. 49;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKORLKYAOEE 20  
||| ||| :||| :|||  
Db 694 POSTLTHVGKRIKLRPEE 713  
RESULT 20  
PARC\_SALTY STANDARD; PRT; 752 AA.  
AC P26973;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Topoisomerase IV subunit A (EC 5.99.1.-).  
GN PARC OR STM3174.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LT2;





RT pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE-20330349; PubMed-10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 Shiba T., Ishii K., Hattori M., Kohara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CW029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON  
 CC GRADIENT ACROSS THE MEMBRANE.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) -> ADP + phosphate +  
 CC H(+)(out).  
 CC -1- SIMILARITY: BELONGS TO THE V-ATPASE D SUBUNIT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE001594; AAD18243.1; -  
 DR EMBL: AE002227; AAF73696.1; -  
 DR EMBL: AP002545; BAA98300.1; -  
 DR TIGR: CP0684; -  
 DR InterPro: IPR002699; ATP-synt\_D.  
 DR Pfam: PF01813; ATP-synt\_D; 1.  
 DR ProDom: PD004122; ATP-synt\_D; 1.  
 DR Hydrolyse: ATP synthetase; Hydrogen ion transport; Complete proteome.  
 KW SEQUENCE 209 AA; 2378 MW; D40A5C36313FE6B CRC64;  
 SQ

Query Match 44.1%; Score 45; DB 1; Length 209;  
 Best Local Similarity 50.0%; Pred. No. 4.2;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 KEIRLEVGKQRLKVAQEE 20  
 ID 1: |||:||||:  
 DB 133 KRYMAVSKERLKLIEE 150

RESULT 13  
 ILVE\_RICPR STANDARD; PRT; 290 AA.  
 AC 005970:  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)  
 GN (BCAT).  
 GN ILVE OR RP428  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE-97419517; PubMed-9274032;  
 RA Andersson J.O., Andersson S.G.E.;  
 RT "Genomic rearrangements during evolution of the obligate  
 RT intracellular parasite Rickettsia prowazekii as inferred from an  
 RT analysis of 52015 bp nucleotide sequence."  
 RL Microbiology 143:2783-2795(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE-99039499; PubMed-9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria."  
 RL Nature 396:133-140(1998).  
 CC -1- CATALYTIC ACTIVITY: L-LEUCINE + 2-OXOGUTARATE -> 4-METHYL-2-  
 CC OXOPENTANATE + L-GLUTAMATE (ALSO ACTS ON L-ISOLEUCINE AND  
 CC L-VALINE).  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Y11777; CAA72450.1; -  
 DR EMBL: AJ235271; CAA14885.1; -  
 DR HSSP: P00510; 1A3G.  
 DR InterPro: IPR001544; AminoTran\_4.  
 DR Pfam: PF01063; AminoTran\_4; 1.  
 DR ProDom: PD001961; AminoTran\_4; 1.  
 DR PROSITE: PS00770; AA\_TRANSFERRER\_CLASS\_4; 1.  
 DR Transferrase; AminoTransferase; Branched-chain amino acid biosynthesis;  
 KW Pyridoxal phosphate; Complete proteome.  
 FT BINDING 155 155 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 290 AA; 33068 MW; 98374E325350763D CRC64;

Query Match 43.1%; Score 44; DB 1; Length 290;  
 Best Local Similarity 61.1%; Pred. No. 8.6;  
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 KEIRLEVGKQRLKVAQEE 20  
 ID 1: |||:||||:  
 DB 225 KSLCLEVKEERLKLQIE 242

RESULT 14  
 FOJO\_DROME STANDARD; PRT; 583 AA.  
 AC P54360; Q24176;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Four-jointed protein.  
 GN FJ.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Eye imaginal disk;  
 RX MEDLINE-96038089; PubMed-7555705;  
 RA Villano J.L., Katz F.N.;  
 RT "four-jointed is required for intermediate growth in the proximal-  
 RT distal axis in Drosophila."  
 RL Development 121:2767-2777(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE-96187865; PubMed-8606003;  
 RA Brodsky M.H., Steller H.;  
 RT "Positional information along the dorsal-ventral axis of the  
 RT Drosophila eye: graded expression of the four-jointed gene."  
 RL Dev. Biol. 173:428-446(1996).

RA MEDLINE-91311420; PubMed-1649901;  
RA Tsarev S.A., Emerson S.U., Balaayan M.S., Ticehurst J.R.,  
RA Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
RT structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE-99232168; PubMed-2541023;  
RA Balaayan M.S., Kusov Y.Y., Andjapardze A.G., Tsarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vassilenco S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
RT and simian hepatitis A viruses.";  
RL FEBS Lett. 247:425-428(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D00924; BAA00766.1; -;  
DR EMBL: X15461; CAA33490.1; -;  
DR PIR: A30470; GNNYSA.  
DR PIR: S04885; S04885.  
DR MEROPS: C03.005; -;  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam: PF00910; RNA\_helicase\_1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 27  
FT CHAIN 28 249  
FT CHAIN 250 495  
FT CHAIN 496 795  
FT CHAIN 796 984  
FT CHAIN 985 1091  
FT CHAIN 1092 1426  
FT CHAIN 1427 1498  
FT CHAIN 1499 1521  
FT CHAIN 1522 1741  
FT CHAIN 1742 2230  
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;  
  
Query Match 93.1%; Score 95; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 5.6e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 PYKELRELVGKORLKYAOEE 20  
Db 803 PYKELRMEVKGKORLKYAMEE 822  
|||||  
RESULT 11  
POLG\_HPAVT STANDARD; PRT; 839 AA.  
AC P31788;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
DE P2a] (Fragment).  
OS Simian hepatitis A virus (strain CY-145).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.  
OX NCBI\_TaxID=31707;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91311421; PubMed-1649902;  
RA Nainan O.V., Margolis H.S., Robertson B.H., Balaayan M., Brinton M.A.;  
RT "Sequence analysis of a new hepatitis A virus naturally infecting  
RT cynomolgus macaques (Macaca fascicularis).";  
RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M59286; AAA45473.1; -;  
DR PIR: J01180; GNNY82.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 ?  
FT CHAIN ? >839  
FT CHAIN 839 839  
SQ SEQUENCE 839 AA; 93825 MW; 2CACC4BD1E192DBC CRC64;  
  
Query Match 92.2%; Score 94; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 2.9e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 PYKELRELVGKORLKYAOEE 20  
Db 798 PYKELRLEVGKORLKYAREE 817  
|||||  
RESULT 12  
VATD\_CHLPN STANDARD; PRT; 209 AA.  
AC G92991;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE V-type ATP synthase subunit D (EC 3.6.3.14) (V-type ATPase subunit D).  
GN ATPD OR CPN0090 OR CP0684.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83536;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CWL029;  
RX MEDLINE-99206606; PubMed-10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger U., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AR39;  
RX MEDLINE-20150255; PubMed-10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA Linhe O., Hickey E.K., Peterson J., Uiterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Eskin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia

16-OCT-2001 (Rel. 40, last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
RX NCBI\_TaxID=12099;  
OK  
RN  
RP SEQUENCE FROM N.A.  
RA MEDLINE=85190549; PubMed=2986127;  
RA Natarajan R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -1- SUBUNTT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PMM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: K02990; AAA45472.1; -  
DR PIR: A03903; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1076  
FT CHAIN 1077 1422  
FT CHAIN 1423 1484  
FT CHAIN 1485 1507  
FT CHAIN 1508 1678  
FT CHAIN 1679 2227  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;  
  
Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 PYKELRLEVGKORLKYAOEE 20  
DB 799 PYKELRLEVGKORLKYAOEE 818  
  
RESULT 9  
POLG\_HPAVM STANDARD: PRT; 2227 AA.  
AC P13901; 081083; 081084; 081085; 081086; 081087; 081088; 081089;  
AC 081090; 081091; 081092; 081093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MBB).  
CC  
RN  
RP SEQUENCE FROM N.A.

Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OK NCBI\_TaxID=12100;  
RX NCBI\_TaxID=12100;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=86045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wissel T., Klein R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
RT virus (isolate MBB).";  
RL Virus Res. 8:153-171(1987).  
CC -1- SUBUNTT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PMM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: M20273; AAA45474.1; -  
DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
  
Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 PYKELRLEVGKORLKYAOEE 20  
DB 799 PYKELRLEVGKORLKYAOEE 818  
  
RESULT 10  
POLG\_HPAVS STANDARD: PRT; 2230 AA.  
AC P14553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OK NCBI\_TaxID=12102;  
RN  
RP SEQUENCE FROM N.A.



CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 CC VP3, AND VP4.  
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -1- PSM:LADEOS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
 CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
 CC SHOWN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/1sb.ch>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M14114; AAA45475.1; -;  
 CC EMBL; M14707; AAA45465.1; -;  
 CC EMBL; M14707; AAA45466.1; ALT\_INIT.  
 CC EMBL; M16632; AAA45471.1; -;  
 CC PIR; A25981; GNNYHM.  
 CC PIR; A25914; GNNYMK.  
 CC PIR; A03905; A03905.  
 CC MEROPS; C03.005; -;  
 CC InterPro: IPR000605; RNA\_helicase.  
 CC InterPro: IPR001205; RNA\_pol\_P3D.  
 CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam; PF00910; RNA\_helicase; 1.  
 CC KW Polypeptide; Coat protein; Core protein; Hydrolase; Thiol protease.  
 CC RNA-directed RNA polymerase;  
 CC FT CHAIN 1 23  
 CC FT CHAIN 24 245  
 CC FT CHAIN 246 491  
 CC FT CHAIN 492 836  
 CC FT CHAIN 837 980  
 CC FT CHAIN 981 1087  
 CC FT CHAIN 1088 1422  
 CC FT CHAIN 1423 1496  
 CC FT CHAIN 1497 1519  
 CC FT CHAIN 1520 1738  
 CC FT CHAIN 1739 2227  
 CC FT CHAIN 77 77  
 CC FT VARIANT 764 764  
 CC FT VARIANT 821 821  
 CC FT VARIANT 1052 1052  
 CC FT VARIANT 1062 1062  
 CC FT VARIANT 1118 1118  
 CC FT VARIANT 1151 1151  
 CC FT VARIANT 1163 1163  
 CC FT VARIANT 1277 1277  
 CC FT VARIANT 1500 1500  
 CC FT VARIANT 1805 1805  
 CC FT VARIANT 1930 1930  
 CC FT VARIANT 2227 AA; 251506 MM; 01E225E7AE8740A6 CRC64;  
 CC  
 CC RNA-DIRECTED POLYMERASE 3D.  
 CC K -> R (IN ATTENUATED STRAIN).  
 CC E -> V (IN ATTENUATED STRAIN).  
 CC N -> S (IN ATTENUATED STRAIN).  
 CC A -> V (IN ATTENUATED STRAIN).  
 CC N -> S (IN ATTENUATED STRAIN).  
 CC G -> A (IN ATTENUATED STRAIN).  
 CC K -> M (IN ATTENUATED STRAIN).  
 CC E -> K (IN ATTENUATED STRAIN).  
 CC F -> S (IN ATTENUATED STRAIN).  
 CC V -> I (IN ATTENUATED STRAIN).  
 CC H -> Y (IN ATTENUATED STRAIN).  
 CC D -> N (IN ATTENUATED STRAIN).  
 CC S -> T (IN ATTENUATED STRAIN).  
 CC  
 CC QUERY MATCH 100.0%; Score 102; DA 1; Length 2227;  
 CC Best Local Similarity 100.0%; Pred. No. 4,3e-08;  
 CC Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 PYKELRLVGVKORUKYAOEE 20  
 CC ||||||||||||||||||||  
 CC Db 799 PYKELRLVGVKORUKYAOEE 818  
 CC  
 CC RESULT 8  
 CC ID POLG\_HPAVL STANDARD; PRT; 2227 AA.  
 CC AC P06441;  
 CC DT 01-JAN-1988 (Rel. 06, Created)  
 CC DT 01-JAN-1988 (Rel. 06, Last sequence update)

```
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M59810; AAA45468.1; -.
CC
CC MEROPS: C03.005; -.
CC
CC InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol.1.
CC
CC Pfam: PF00910; RNA_helicase.1.
CC
CC Polyprotein: Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC
CC CHAIN 1 23
CC CHAIN 24 245
CC CHAIN 246 491
CC CHAIN 492 794
CC CHAIN 795 900
CC CHAIN 901 1087
CC CHAIN 1088 1422
CC CHAIN 1423 1495
CC CHAIN 1496 1518
CC CHAIN 1519 1737
CC CHAIN 1738 2226
CC CHAIN 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;
CC
CC SEQUENCE
```

```
Query Match 100.0%; Score 102; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PYKELRLEVGKORLKXAOEE 20
DB 799 PYKELRLEVGKORLKXAOEE 818
```

```
RESULT 5
POLG_HPAV4 STANDARD; PRT; 2226 AA.
ID POLG_HPAV4
AC P26581;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 43c).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_Taxid=12095;
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA "Antigenic and genetic variation in cytopathic hepatitis A virus
RA variants arising during persistent infection: evidence for genetic
RA recombination."
RT J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M59809; AAA45469.1; -.
CC
CC MEROPS: C03.005; -.
CC
CC InterPro: IPR001205; RNA_helicase.
CC Pfam: PF00680; RNA_dep_RNA_pol.1.
CC
CC Pfam: PF00910; RNA_helicase.1.
CC
CC Polyprotein: Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC
CC CHAIN 1 23
CC CHAIN 24 245
CC CHAIN 246 491
CC CHAIN 492 794
CC CHAIN 795 900
CC CHAIN 901 1087
CC CHAIN 1088 1422
CC CHAIN 1423 1495
CC CHAIN 1496 1518
CC CHAIN 1519 1737
CC CHAIN 1738 2226
CC CHAIN 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;
CC
CC SEQUENCE
```

```
Query Match 100.0%; Score 102; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PYKELRLEVGKORLKXAOEE 20
DB 799 PYKELRLEVGKORLKXAOEE 818
```

```
RESULT 6
POLG_HPAV8 STANDARD; PRT; 2226 AA.
ID POLG_HPAV8
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 18f).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_Taxid=12096;
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA "Antigenic and genetic variation in cytopathic hepatitis A virus
RA variants arising during persistent infection: evidence for genetic
RA recombination."
RT J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

S0 SEQUENCE 341 AA; 38003 MW; 066918289BFI2605 CMC64;

Query Match 100.0%; Score 102; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 5,9e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLVGVKORLKYAOEE 20
   |||||||
Db 303 PYKELRLVGVKORLKYAOEE 322

RESULT 2
POLG_HPAAVC STANDARD; PRT; 808 AA.
AC Q02381;
DT 01-JUL-1993 (Rel. 26, Created)
DR 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Coat proteins VP1 70 VP4; Core protein
DE P2A] (Fragment).
OS Hepatitis A virus (strain GAT5).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=31706;
[]
RP SEQUENCE FROM N.A.
RA MEDLINE=92260183; PubMed=1316423;
RA Khanna B., Speldring J.E., Innis B.L., Robertson B.H.;
RT "Characterization of a genetic variant of human hepatitis A virus.";
J. Med. Virol. 36:118-124(1992)
-1 SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
-1 PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CR EMBL: M66695; AAA45477.1; -
CW Polypeptide; Coat protein; Core protein.
KM POLYPROTEIN; 1
FT NON_TER 1
FT CHAIN <1 2 COAT PROTEIN VP4 (PIA).
FT CHAIN 3 223 COAT PROTEIN VP2 (PIB).
FT CHAIN 224 470 COAT PROTEIN VP3 (PIC).
FT CHAIN 471 770 COAT PROTEIN VP1 (PID).
FT CHAIN 771 >808 CORE PROTEIN P2A.
FT NON_TER 808
SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 100.0%; Score 102; DB 1; Length 808;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PYKELRLVGVKORLKYAOEE 20
   |||||||
Db 778 PYKELRLVGVKORLKYAOEE 797

RESULT 3
POLG_HPAAVC STANDARD; PRT; 852 AA.
AC P06442; O83741; O83742;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Coat proteins VP1 70 VP4; Core protein

```

```

DE P2A1 (Fragment).
OS Hepatitis A virus (strain CR326).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_Taxid=12097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85185648; PubMed=2985793;
RA Lihemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,
RA Young A., Milra S.W.;
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA.";
RL J. Virol. 54:247-255(1985).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M10033; AAA45470.1; -.
DR PIR; A03904; GNNYHA.
KW Polypeptide; Coat protein; Core protein.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).
FT CHAIN 837 >852 CORE PROTEIN P2A.
FT NON_TER 852 852
SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;

Query Match 100.0%; Score 102; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 1,6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELLEVGRKRLKYAOEE 20
Db 799 PYKELLEVGRKRLKYAOEE 818

RESULT 4
POLG_HPAV2 STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_Taxid=12094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination.";
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:10:05 ; Search time 59.43 Seconds  
(without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432A-40  
Perfect score: 102  
Sequence: I PYKELRLVGGKQRLKYAQEE 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : SwissProt\_40:\*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	102	100.0	341	1	POLG_HPVAV1	P13672 hepatitis a
2	102	100.0	808	1	POLG_HPVAV3	P002381 hepatitis a
3	102	100.0	852	1	POLG_HPVAV2	P06442 hepatitis a
4	102	100.0	2226	1	POLG_HPVAV2	P26580 hepatitis a
5	102	100.0	2226	1	POLG_HPVAV4	P26581 hepatitis a
6	102	100.0	2226	1	POLG_HPVAV8	P25862 hepatitis a
7	102	100.0	2227	1	POLG_HPVAV4	P08617 hepatitis a
8	102	100.0	2227	1	POLG_HPVAVL	P06441 hepatitis a
9	102	100.0	2227	1	POLG_HPVAVM	P13901 hepatitis a
10	95	93.1	2230	1	POLG_HPVAV5	P14553 simian hepa
11	94	92.2	839	1	POLG_HPVAVT	P11788 simian hepa
12	45	44.1	209	1	VATD_CHLPPN	O92991 chlamydia f
13	44	43.1	290	1	ILVE_RITCRP	O05970 rickettsia
14	44	43.1	583	1	FOLQ_DRITRME	P54360 drosophila
15	43	42.2	208	1	UPP_AQUAE	P067914 aquifex aec
16	43	42.2	480	1	DNNA_RHITIME	P35890 rhizobium n
17	43	42.2	179	1	VEGL_ECOLI	O30408 b. thurocidat
18	42.5	41.7	3587	1	TYCR_BACBR	P76573 escherichia
19	42	41.2	752	1	PARC_ECOLI	P20082 escherichia
20	42	41.2	752	1	PARC_ECOLI	P26973 salmonella
21	42	41.2	939	1	SYV_CHLTR	O84304 chlamydia t
22	41	40.2	68	1	GBG5_HUMAN	P30670 homo sapien
23	41	40.2	68	1	GBG5_HUMAN	P50151 homo sapien
24	41	40.2	147	1	GST2_HUMAN	O99723 homo sapien
25	41	40.2	318	1	VZ34_AQUAE	O66423 aquifex aec
26	41	40.2	446	1	ATPD_MITCU	Q10594 mycobacteri
27	40	39.2	69	1	YOP_HAETIN	P45183 haemophilus
28	40	39.2	353	1	V070_ARCFU	O30166 archaeoglobi
29	40	39.2	456	1	BEFL_DIDI	P18624 dictyostelm
30	40	39.2	470	1	BEFL_YEAST	P38994 saccharomy
31	40	39.2	685	1	STMI_MOUSE	O13566 homo sapien
32	40	39.2	685	1	STMI_MOUSE	P70302 homo muscul
33	40	39.2	723	1	SYM_PYRIO	O587721 pyrococcus

34	40	39.2	73.3	1	ERG7_RAT	p17480	rattus norvegicus
35	40	39.2	76.1	1	UBF1_HUMAN	p17480	homo sapiens
36	40	39.2	76.4	1	UBF1_RAT	p25977	rattus norvegicus
37	40	39.2	76.5	1	UBF1_MOUSE	p25977	mus musculus
38	40	39.2	79.5	1	ION_MYCPN	p78025	mycoplasma
39	40	39.2	93.9	1	SVY_CHILM	g93k01	chlamydia m
40	40	39.2	38.59	1	RPOD_LELV	004561	lelystady m
41	39.5	38.7	97.1	1	SN2L_CAEBL	PA1877	caenorhabditis
42	39.5	38.7	97.6	1	SN2L_HUMAN	PA8370	homo sapiens
43	39.5	38.7	181.2	1	BRC1_MOUSE	PA8754	mus musculus
44	39	38.2	46	1	DIUW_LOCOM	323465	locusta migratoria
45	39	38.2	36.2	1	LIVE_STRCO	068650	streptococcus
46	39	38.2	40.2	1	APL3_HUMAN	095226	homo sapiens
47	39	38.2	42.4	1	EXON_NPVOP	p24001	oryzias latipes
48	39	38.2	50.8	1	V56K_PLRVL	p17525	potato lea
49	39	38.2	50.8	1	V56K_PLRWP	p11626	potato lea
50	39	38.2	65.3	1	PABP_SCHPO	p31209	schizosaccharomyces
51	39	38.2	14.77	1	REST_HUMAN	P30622	homo sapiens
52	39	38.2	18.88	1	TRHY_HUMAN	007283	homo sapiens
53	39	38.2	19.01	1	TCF1_TOBAC	P12222	nicotiana glauca
54	38.5	37.7	41.6	1	RP5D_MICAE	P52332	microcystis aeruginosa
55	38.5	37.7	64.1	1	DNAM_METSS	09z636	methylorubrum
56	38	37.3	18.6	1	TNR2_ECOLI	PA4130	escherichia coli
57	38	37.3	22.6	1	PIRM_MOUSE	P33506	mus musculus
58	38	37.3	26.4	1	RPOD_SULAC	P39471	sulfolobus solfataricus
59	38	37.3	27.5	1	MADE_ECOLI	PA1883	escherichia coli
60	38	37.3	28.9	1	SGOD_MESAU	p97201	mesocricetus auratus

## ALIGNMENTS

RESULT 1			
POLG_HPAV1	STANDARD:	PRT:	341 AA.
ID	POLG_HPAV1		
AC	P13672;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Genome polypeptide [Contains: Coat proteins VP1 TO VP3; Core protein P2a] (Fragment).		
OS	Hepatitis A virus (strain LCDC-1).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;		
OC	Hepatovirus.		
OX	NCBI_TaxID=12093;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89263805; Pubmed=2542903;		
RA	Andonov A.P., Lau P., Chaudhary R.;		
RT	"Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus (HAV)."		
RL	Nucleic Acids Res. 17:3594-3594(1989).		
CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,		
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,		
CC	VP3, AND VP4.		
CC	-1- PWM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD NATURE PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ,		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL, X14666; CAA32794.1; .		
DR	PIR: S04137; S04137.		
KW	Polypeptide; Coat protein; Core protein.		
FT	NON_TER 1 1		
FT	CHAIN <1 1 COAT PROTEIN VP3 (1C).		
FT	CHAIN 2 340 COAT PROTEIN VP1 (1D).		
FT	CHAIN 341 >341 CORE PROTEIN P2A.		
FT	NON_TER 341 341		

**THIS PAGE BLANK (USPTO)**



Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKLRLVKGKORLKYAOE 20  
| | | | | | | | | |  
Db 575 KRLQERREORLKRQEE 592

RESULT 24  
US-08-800-644-94  
; Sequence 94, Application US/08800644  
; Patent No. 5958752  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; City: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,644  
; FILING DATE: 14-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/056,200  
; FILING DATE: 30-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-800-644-94

Query Match 38.28; Score 39; DB 2; Length 1898;  
Best Local Similarly 55.68; Pred. NO. 3.9e+02;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKLRLVKGKORLKYAOE 20  
| | | | | | | | | |  
Db 575 KRLQERREORLKRQEE 592

RESULT 25  
US-08-961-083-160  
; Sequence 160 Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 160:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 641 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-961-083-160

Query Match 37.78; Score 38.5; DB 4; Length 641;  
Best Local Similarly 55.68; Pred. NO. 1.5e+02;  
Matches 10; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 2 YKRLRLVKGKORLKYAOE 19  
| | | | | | | | | |  
Db 85 YKEYR-EVQNRSKYKSD 101

Search completed: June 16, 2002, 00:03:14  
Job time: 8528 sec

```

ZIP: 84111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS95
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,863
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/157,005
FILING DATE: 26-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91201398.4
FILING DATE: 06-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 92200781.0
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL92/00096
FILING DATE: 05-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Turner, Allen C.
REGISTRATION NUMBER: 33041
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801) 532-1922
TELEFAX: (801) 531-9168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-747-863-3

Query Match 39.2%; Score 40; DB 4; Length 1463;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PYKELRLLEVQKORLKYAQ 18
| | | | | | | | | | | | | | | | | | | |
Db 1444 PGTQLVQLGKRLPPGQ 1461

RESULT 22
US-08-612-521-2
Sequence 2, Application US/08612521
Patent No. 5786463
GENERAL INFORMATION:
APPLICANT: Peery, Robert B
APPLICANT: Skatrud, Paul L
APPLICANT: Thornevell, Susan J
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patent Division/AEH
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,521
FILING DATE:

```

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Hamilton, Amy E
  REGISTRATION NUMBER: 33,894
  REFERENCE/DOCKET NUMBER: X-9693
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 317-276-3169
    TELEFAX: 317-276-1294
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1408 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
    MOLECULE TYPE: protein
  US-08-612-521-2

Query Match          38.2%; Score 39; DB 1; Length 1408;
Best Local Similarity 40.0%; Pred. No. 2.9e+02;
Matches      8; Conservative      4; Mismatches      8; Indels      0; Gaps      0.

QY      1 PYKELRLVGRQRLKYAEE 20
      |||:::| | |::||
      104 PYKSKFDFLKSRRKKEEE 123

RESULT 23
US-08-056-200-94
Sequence 94, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
  APPLICANT: Steinhert, Peter M.
  APPLICANT: Lee, Seung-Chul
  APPLICANT: Kim, In-Gyu
  APPLICANT: Chung, Soo-Il
  APPLICANT: Park, Sang-Chul
  TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
  NUMBER OF SEQUENCES: 117
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Knobbe, Martens, Olson & Bear
    STREET: 620 Newport Center Drive, Sixteenth Floor
    CITY: Newport Beach
    STATE: CA
    COUNTRY: U.S.A.
    ZIP: 92660
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentln Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/056,200
    FILING DATE: 30-APR-1993
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Fedtich, Michael F.
    REGISTRATION NUMBER: 36,799
    REFERENCE/DOCKET NUMBER: NH054.001A
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: (714) 760-0404
      TELEFAX: (714) 760-9502
    INFORMATION FOR SEQ ID NO: 94:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1898 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
      MOLECULE TYPE: protein
    US-08-056-200-94

Query Match          38.2%; Score 39; DB 1; Length 1898;
Best Local Similarity 55.6%; Pred. No. 3.9e+02;

```



RESULT 19  
US-09-080-643-4  
; Sequence 4, Application US/09080643  
; Patent No. 6303771  
; GENERAL INFORMATION:  
; APPLICANT: Zalacain, Magdalena  
; APPLICANT: Biswas, Sanjoy  
; APPLICANT: Ingraham, Karen A.  
; APPLICANT: Chalker, Allison F.  
; APPLICANT: Holmes, David J.  
; APPLICANT: Traint, Christopher M.  
; APPLICANT: Warren, Richard L.  
; APPLICANT: Brown, James R.  
; APPLICANT: Throup, John P.  
; APPLICANT: Laylor, Elizabeth J.  
; TITLE OF INVENTION: pth  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,643  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION NUMBER:  
; FILING DATE: 20-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10117  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 189 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-080-643-4

Query Match 39.2%; Score 40; DB 4; Length 189;  
Best Local Similarity 57.1%; Pred. NO. 24;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 YKELRLVGVKORLK 15  
Db 92 YDLDMEVGRIRLR 105

RESULT 20  
US-08-157-005-3  
; Sequence 3, Application US/08157005  
; Patent No. 5620691  
; GENERAL INFORMATION:  
; APPLICANT: Wensvoort, Gert  
; APPLICANT: Terpstra, Catharinus  
; APPLICANT: Pol, Johannes M  
; APPLICANT: Moorman, Robertus J  
; APPLICANT: Meulenber, Johanna J

; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,  
; TITLE OF INVENTION: VACCINE COMPOSITIONS AND DIAGNOSTIC KITS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/157,005  
; FILING DATE: 26-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91201398.4  
; FILING DATE: 06-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92200781.0  
; FILING DATE: 18-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/NL92/00096  
; FILING DATE: 05-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Moran, Thomas F  
; REGISTRATION NUMBER: 16,579  
; REFERENCE/DOCKET NUMBER: 44819  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 315-1931  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1463 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-157-005-3

Query Match 39.2%; Score 40; DB 1; Length 1463;  
Best Local Similarity 50.0%; Pred. NO. 2.1e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKORLK 18  
Db 1444 PGTLOVELGKRPQ 1461

RESULT 21  
US-08-747-863-3  
; Sequence 3, Application US/08747863  
; Patent No. 6197310  
; GENERAL INFORMATION:  
; APPLICANT: Wensvoort, Gert  
; APPLICANT: Terpstra, Catharinus  
; APPLICANT: Pol, Johannes M  
; APPLICANT: Moorman, Robertus J  
; APPLICANT: Meulenber, Johanna J  
; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,  
; TITLE OF INVENTION: VACCINE COMPOSITIONS AND DIAGNOSTIC KITS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt & Rossa  
; STREET: 525 South 300 East  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0411 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 147 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: COLNOT07  
CLONE: 903729  
US-08-959-865-3

Query Match 40.2%; Score 41; DB 2; Length 147;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 LEVGKORLKY 16  
|:|:|:|:|:|:|  
Db 24 LQVGKARLKY 33

RESULT 17  
US-08-842-234-2  
Sequence 2, Application US/08842234  
Patent No. 6147050  
GENERAL INFORMATION:  
APPLICANT: Gentz, Reiner  
APPLICANT: Fleischmann, Robert D.  
TITLE OF INVENTION: 5-LIPOXYGENASE-ACTIVATING PROTEIN II  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CROCHI,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NJ  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/842,234  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/264,003  
FILING DATE: 06-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-117  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 147 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-842-234-2

Query Match 40.2%; Score 41; DB 4; Length 147;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 LEVGKORLKY 16  
|:|:|:|:|:|:|  
Db 24 LQVGKARLKY 33

RESULT 18  
US-09-080-643-2  
Sequence 2, Application US/09080643  
Patent No. 6303771  
GENERAL INFORMATION:  
APPLICANT: Zalacain, Magdalena  
APPLICANT: Biswas, Sanjoy  
APPLICANT: Ingraham, Karen A.  
APPLICANT: Chalker, Alison F.  
APPLICANT: Holmes, David J.  
APPLICANT: Traini, Christopher M.  
APPLICANT: Warren, Richard L.  
APPLICANT: Brown, James R.  
APPLICANT: Throup, John P.  
APPLICANT: Lawlor, Elizabeth J.  
TITLE OF INVENTION: Plh  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/080,643  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/066,998  
FILING DATE: 20-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GM10117  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 189 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-080-643-2

Query Match 39.2%; Score 40; DB 4; Length 189;  
Best Local Similarity 57.1%; Pred. No. 24;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 YKELRLVYGKORLK 15  
|:|:|:|:|:|:|  
Db 92 YDDLDMVEYKIRLR 105

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
PCT-US95-06406A-5

Query Match          40.2%; Score 41; DB 5; Length 69;
Best Local Similarity 50.0%; Pred. No. 5.7;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 KELRLVGGKRLRYA 18
   :|:|:|:|:|:|:|:|
Db 15 QQLRLVGLNRYKVSQ 30

RESULT 14
PCT-US95-06406A-7
; Sequence 7, Application PC/TUS9506406A
; GENERAL INFORMATION:
; APPLICANT: Janet D. Robshaw, Charles Kunsch
; TITLE OF INVENTION: CDNA Clones Encoding Human G Protein
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
PCT-US95-06406A-7

Query Match          40.2%; Score 41; DB 5; Length 69;
Best Local Similarity 43.8%; Pred. No. 5.7;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 KELRLVGGKRLRYA 18
   :|:|:|:|:|:|:|:|
Db 15 QQLRLVGLNRYKVSQ 30

RESULT 15
US-08-264-003B-2
; Sequence 2, Application US/08264003B
; Patent No. 5696076
; GENERAL INFORMATION:
```

```

; APPLICANT: Gentz, Relner
; APPLICANT: Fleischmann, Robert D.
; TITLE OF INVENTION: 5-LIPOXYGENASE-ACTIVATING PROTEIN II
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,003B
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-117
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 147 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-264-003B-2

Query Match          40.2%; Score 41; DB 1; Length 147;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 7 LEVGRKRLKY 16
   :|:|:|:|:|:|:|
Db 24 LQVGRKRLKY 33

RESULT 16
US-08-959-865-3
; Sequence 3, Application US/08959865
; Patent No. 5919627
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: MICROSOMAL GLUTATHIONE-S TRANSFERASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,865
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/111,348  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/606,789  
FILING DATE:  
APPLICATION NUMBER: 08/440,743  
FILING DATE: May 5, 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/320,011  
FILING DATE: October 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Luther, Barbara J.  
REGISTRATION NUMBER: 33,954  
REFERENCE/DOCKET NUMBER: PF-0055 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-852-0195  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: PITUITARY  
CLONE: 112530  
US-09-111-348-2

Query Match 40.2%; Score 41; DB 2; Length 68;  
Best Local Similarity 50.0%; Pred. No. 5.6;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KRLREVGKORLKVQ 18  
DB 15 QQLRLAGLNKRVKVSQ 30

RESULT 12  
US-09-111-348-4  
Sequence 4, Application US/09111348  
Patent No. 5912130  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Stuart, Susan G.  
APPLICANT: Murry, Lynn E.  
APPLICANT: Guegler, Kirk J.  
APPLICANT: Seihamer, Jeffrey J.  
TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
TITLE OF INVENTION: GAMMA-5 SUBUNIT  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/111,348  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/606,789  
FILING DATE:  
APPLICATION NUMBER: 08/440,743  
FILING DATE: May 5, 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/320,011  
FILING DATE: October 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Luther, Barbara J.  
REGISTRATION NUMBER: 33,954  
REFERENCE/DOCKET NUMBER: PF-0055 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-852-0195  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 204241  
US-09-111-348-4

Query Match 40.2%; Score 41; DB 2; Length 68;  
Best Local Similarity 50.0%; Pred. No. 5.6;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KRLREVGKORLKVQ 18  
DB 15 QQLRLAGLNKRVKVSQ 30

RESULT 13  
PCT-US95-06406A-5  
Sequence 5, Application PC/TUS9506406A  
GENERAL INFORMATION:  
APPLICANT: Janet D. Robishaw, Charles Kunsch  
TITLE OF INVENTION: CDNA Clones Encoding Human G Protein  
TITLE OF INVENTION: Subunits  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE:  
STREET:  
CITY:  
STATE:  
COUNTRY:  
ZIP:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06406A  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:

APPLICANT: Stuart, Susan G.  
APPLICANT: Murry, Lynn E.  
APPLICANT: Guegler, Karik J.  
APPLICANT: Selhamer, Jeffrey J.  
TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
TITLE OF INVENTION: GAMMA-5 SUBUNIT  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/606,789  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,743  
FILING DATE: May 5, 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/320,011  
FILING DATE: October 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Luther, Barbara J.  
REGISTRATION NUMBER: 33,954  
REFERENCE/DOCKET NUMBER: PF-0055 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-852-0195  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: PITUITARY  
CLONE: 112530  
US-08-606-789-2

Query Match 40.2%; Score 41; DB 1; Length 68;  
Best Local Similarity 50.0%; Pred. No. 5.6;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KRLRLEVGKRLKRYAQ 18  
Db 15 QQLRLEAGLNRKVSQ 30

RESULT 10  
US-08-606-789-4  
Sequence 4, Application US/08606789  
Patent No. 5783418  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Stuart, Susan G.  
APPLICANT: Murry, Lynn E.  
APPLICANT: Guegler, Karik J.  
APPLICANT: Selhamer, Jeffrey J.  
TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
TITLE OF INVENTION: GAMMA-5 SUBUNIT  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/606,789  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,743  
FILING DATE: May 5, 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/320,011  
FILING DATE: October 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Luther, Barbara J.  
REGISTRATION NUMBER: 33,954  
REFERENCE/DOCKET NUMBER: PF-0055 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-852-0195  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 204241  
US-08-606-789-4

Query Match 40.2%; Score 41; DB 1; Length 68;  
Best Local Similarity 50.0%; Pred. No. 5.6;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KRLRLEVGKRLKRYAQ 18  
Db 15 QQLRLEAGLNRKVSQ 30

RESULT 11  
US-09-111-348-2  
Sequence 2, Application US/09111348  
Patent No. 5912130  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Stuart, Susan G.  
APPLICANT: Murry, Lynn E.  
APPLICANT: Guegler, Karik J.  
APPLICANT: Selhamer, Jeffrey J.  
TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
TITLE OF INVENTION: GAMMA-5 SUBUNIT  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262051  
CURRENT APPLICATION NUMBER: US/08/397,232A  
CURRENT FILING DATE: 1995-04-17  
EARLIER APPLICATION NUMBER: 07/947,338  
EARLIER FILING DATE: 1992-09-18  
EARLIER APPLICATION NUMBER: PCT/US93/08610  
EARLIER FILING DATE: 1993-09-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO: 4  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 102; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 799 PYKELRLEVGKORLKYAOEE 818

RESULT 7  
US-09-171-387-2  
Sequence 2, Application US/09171387  
Patent No. 6280734  
GENERAL INFORMATION:  
APPLICANT: RAYCHAUDHURI, GOPA;  
EMERSON, SUZANNE, U. ;  
PURCELL, ROBERT, H. ;  
TITLE OF INVENTION: SIMIAN-HUMAN HAV  
HAVING A CHIMERIC 2C PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,387  
FILING DATE: 24-Mar-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/06506  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US60/015,642  
FILING DATE: 19-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feller  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4229051  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2227 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 102; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 799 PYKELRLEVGKORLKYAOEE 818

RESULT 8

US-08-087-016-2  
Sequence 2, Application US/08087016  
Patent No. 5430135

GENERAL INFORMATION:  
APPLICANT: NAINAN, OMANA V.  
APPLICANT: MARGOLIS, HAROLD S.  
APPLICANT: ROBERTSON, BETTY H.  
APPLICANT: BRINTON, MARGO H.  
APPLICANT: EBERT, JAMES W.  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L Street N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,016  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,828  
FILING DATE: 03-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 839 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 92.2%; Score 94; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 1.6e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 798 PYKELRLEVGKORLKYAREE 817

RESULT 9

US-08-606-789-2  
Sequence 2, Application US/08606789  
Patent No. 5783418

GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice

```
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2
```

```
Query Match          100.0%; Score 102; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 PYKELRLEVKGKORLKYAOEE 20
        |||
Db      799 PYKELRLEVKGKORLKYAOEE 818
```

```
RESULT      3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4
```

```
Query Match          100.0%; Score 102; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 PYKELRLEVKGKORLKYAOEE 20
        |||
Db      799 PYKELRLEVKGKORLKYAOEE 818
```

```
RESULT      4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
```

```
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6
```

```
Query Match          100.0%; Score 102; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 PYKELRLEVKGKORLKYAOEE 20
        |||
Db      799 PYKELRLEVKGKORLKYAOEE 818
```

```
RESULT      5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2
```

```
Query Match          100.0%; Score 102; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 PYKELRLEVKGKORLKYAOEE 20
        |||
Db      799 PYKELRLEVKGKORLKYAOEE 818
```

```
RESULT      6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:13 ; Search time 79.04 Seconds

(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-40

Perfect score: 102

Sequence: 1 PYKELRLEVGRKRLKYAOE 20

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	100.0	1091	6 5516630-2	Patent No. 5516630
2	102	100.0	2227	3 US-08-475-886-2	Sequence 2, Appli
3	102	100.0	2227	3 US-08-475-886-4	Sequence 4, Appli
4	102	100.0	2227	3 US-08-475-886-6	Sequence 6, Appli
5	102	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
6	102	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
7	102	100.0	2227	4 US-09-171-387-2	Sequence 2, Appli
8	94	92.2	839	1 US-08-087-016-2	Sequence 2, Appli
9	41	40.2	68	1 US-08-606-789-2	Sequence 2, Appli
10	41	40.2	68	1 US-08-606-789-4	Sequence 4, Appli
11	41	40.2	68	2 US-09-111-348-2	Sequence 2, Appli
12	41	40.2	68	2 US-09-111-348-4	Sequence 4, Appli
13	41	40.2	69	5 PCT-US95-06406A-5	Sequence 5, Appli
14	41	40.2	69	5 PCT-US95-06406A-7	Sequence 7, Appli
15	41	40.2	147	1 US-08-264-003B-2	Sequence 2, Appli
16	41	40.2	147	2 US-08-959-865-3	Sequence 3, Appli
17	41	40.2	147	4 US-08-842-234-2	Sequence 2, Appli
18	40	39.2	189	4 US-09-080-643-2	Sequence 4, Appli
19	40	39.2	189	4 US-09-080-643-4	Sequence 4, Appli
20	40	39.2	1463	1 US-08-157-005-3	Sequence 3, Appli
21	40	39.2	1463	4 US-08-747-863-3	Sequence 3, Appli
22	39	38.2	1408	1 US-08-612-521-2	Sequence 2, Appli
23	39	38.2	1898	1 US-08-056-200-94	Sequence 94, Appli
24	39	38.2	1898	2 US-08-800-644-94	Sequence 160, App
25	38.5	37.7	641	3 US-08-961-083-160	Sequence 8, Appli
26	38	37.3	233	4 US-08-836-236-8	Sequence 2, Appli
27	38	37.3	256	2 US-08-719-758-2	Sequence 2, Appli

28	38	37.3	256	4	US-09-119-827-2	Sequence 2, Appli
29	38	37.3	3072	4	US-09-413-814-93	Sequence 93, Appli
30	38	37.3	3079	4	US-09-413-814-80	Sequence 80, Appli
31	37.5	36.8	18	3	US-08-940-095-229	Sequence 229, App
32	37.5	36.8	18	3	US-08-940-093-229	Sequence 229, App
33	37.5	36.8	18	3	US-08-940-096-229	Sequence 229, App
34	37.5	36.8	18	4	US-09-455-719-229	Sequence 229, App
35	37.5	36.8	18	4	US-09-453-605-229	Sequence 229, App
36	37	36.3	67	3	US-09-120-365-93	Sequence 93, Appli
37	37	36.3	67	3	US-09-515-039-93	Sequence 93, Appli
38	37	36.3	68	1	US-08-606-789-8	Sequence 8, Appli
39	37	36.3	68	2	US-09-111-348-8	Sequence 8, Appli
40	37	36.3	69	5	PCT-US95-06406A-6	Sequence 6, Appli
41	37	36.3	284	2	US-08-766-439-32	Sequence 32, Appli
42	37	36.3	284	2	US-08-766-439-41	Sequence 41, Appli
43	37	36.3	284	2	US-08-766-439-42	Sequence 42, Appli
44	37	36.3	284	2	US-08-766-439-43	Sequence 43, Appli
45	37	36.3	284	2	US-08-766-439-44	Sequence 44, Appli
46	37	36.3	284	2	US-08-766-439-45	Sequence 45, Appli
47	37	36.3	458	1	US-08-336-618-24	Sequence 24, Appli
48	37	36.3	552	3	US-09-120-365-5	Sequence 5, Appli
49	37	36.3	552	4	US-09-515-039-5	Sequence 5, Appli
50	37	36.3	599	3	US-08-556-419-22	Sequence 22, Appli
51	37	36.3	629	3	US-08-556-419-23	Sequence 23, Appli
52	37	36.3	637	2	US-08-428-125-10	Sequence 10, Appli
53	37	36.3	637	2	US-08-455-335-10	Sequence 10, Appli
54	37	36.3	1091	3	US-08-633-768A-2	Sequence 2, Appli
55	36.5	35.8	22	3	US-08-940-095-12	Sequence 12, Appli
56	36.5	35.8	22	3	US-08-940-093-12	Sequence 12, Appli
57	36.5	35.8	22	3	US-08-940-096-12	Sequence 12, Appli
58	36.5	35.8	22	4	US-09-465-719-12	Sequence 12, Appli
59	36.5	35.8	22	4	US-09-453-605-12	Sequence 12, Appli
60	36.5	35.8	345	3	US-09-120-365-73	Sequence 73, Appli

## ALIGNMENTS

RESULT 1  
5516630-2  
Patent No. 5516630  
APPLICANT: TELCEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,  
STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.;  
BAROUDY, BAHIGE M.  
TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/788,262  
FILING DATE: 06-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 256,135  
FILING DATE: 06-OCT-1988  
APPLICATION NUMBER: 654,942  
FILING DATE: 27-SEP-1984  
APPLICATION NUMBER: 537,911  
FILING DATE: 30-SEP-1983  
SEQ ID NO.: 2  
LENGTH: 1091  
5516630-2  
Query Match 100.0%; Score 102; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 PYKELRLEVGRKRLKYAOE 20  
DB 1036 PYKELRLEVGRKRLKYAOE 1055  
RESULT 2  
US-08-475-886-2  
Sequence 2, Application US/08475886A



**THIS PAGE BLANK (USPTO)**



Db 138 kymaevskeri1kilee 155

RESULT 24

ABBS2937

ID ABBS2937 standard; Protein; 211 AA.

AC ABBS2937;

DE 11-FEB-2002 (first entry)

XX Escherichia coli polypeptide SEQ ID NO 1284.

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;

XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

XX systemic infection; non-diarrhoeal infection; septicemia;

XX pyelonephritis; antibiotic resistance.

XX Escherichia coli.

XX WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.

XX 02-FEB-2001; 2001FR-0001449.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Blngen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

XX A library of DNA fragments of Escherichia coli strains for the

XX phylogenetic determination of a given strain comprises polynucleotides of

XX nature B2/D+ A-.

XX Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli

XX strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)

XX and encoded proteins (ABBS2459-ABBS2919) of nature B2/D+A-. The

XX immunosuppressive activity as part of pharmaceutical compositions used to

XX treat, palliate or prevent extra-intestinal E. coli infections. The

XX polypeptides are useful for determining the phylogenetic group of a given

XX E. coli strain. These polypeptides can detect and treat an undesired

XX development of E. coli, particularly an extra-intestinal infection that

XX include systemic and non-diarrhoeal infections such as septicemia,

XX pyelonephritis and meningitis this is particularly advantageous as

XX bacterial resistance is increasing with the more frequent use of broad

XX spectrum antibiotics.

SQ Sequence 211 AA;

Query Match 43.1%; Score 44; DB 22; Length 211;

Best Local Similarity 42.9%; Pred. No. 27;

Matches 9; Conservative 7; Mismatches 3; Indels 2; Gaps 1;

QY 1 PYKEL--RLEVGKQRLKTAOE 19

Db 135 pyqglarevkvprerlkxale 155

RESULT 25

ABBS2971

ID ABBS2971 standard; Protein; 211 AA.

XX ABBS2971;

XX

DT 11-FEB-2002 (first entry)

XX Escherichia coli polypeptide SEQ ID NO 1284.

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;

XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

XX systemic infection; non-diarrhoeal infection; septicemia;

XX pyelonephritis; antibiotic resistance.

XX Escherichia coli.

XX WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.

XX 02-FEB-2001; 2001FR-0001449.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Blngen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

XX A library of DNA fragments of Escherichia coli strains for the

XX phylogenetic determination of a given strain comprises polynucleotides of

XX nature B2/D+ A-.

XX Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli

XX strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)

XX and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature

XX B2/D+A-. The polynucleotides have potential antiinflammatory,

XX antibacterial and immunosuppressive activity as part of pharmaceutical

XX compositions used to treat, palliate or prevent extra-intestinal E. coli

XX infections. The polypeptides are useful for determining the phylogenetic

XX group of a given E. coli strain. These polypeptides can detect and treat

XX an undesired development of E. coli, particularly an extra-intestinal

XX infection that include systemic and non-diarrhoeal infections such as

XX septicemia, pyelonephritis and meningitis this is particularly

XX advantageous as bacterial resistance is increasing with the more

XX frequent use of broad spectrum antibiotics.

SQ Sequence 211 AA;

Query Match 43.1%; Score 44; DB 22; Length 211;

Best Local Similarity 42.9%; Pred. No. 27;

Matches 9; Conservative 7; Mismatches 3; Indels 2; Gaps 1;

QY 1 PYKEL--RLEVGKQRLKTAOE 19

Db 135 pyqglarevkvprerlkxale 155

Search completed: June 16, 2002, 00:01:42  
Job time: 12731 sec

CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Guanine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
XX  
SQ Sequence 26 AA:  
  
Query Match 44.1%; Score 45; DB 22; Length 26;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 12 QRLRYAOEE 20  
|||  
1 qrlkyagee 9  
Db  
  
RESULT 22  
AAV11920  
ID AAV11920 standard; Protein: 124 AA.  
XX  
XX AAV11920;  
XX  
DT 18-JUN-1999 (first entry)  
XX  
XX Human 5' EST secreted protein SEQ ID No: 520.  
XX  
XX Human; secreted protein; EST: expressed sequence tag; diagnosis;  
XX forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
XX upstream regulatory sequence; cytokine activity; cell proliferation;  
XX differentiation; haematopoiesis regulation; tissue growth regulation;  
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; anti-inflammatory; tumour inhibition.  
XX  
XX Homo sapiens.  
XX  
XX WO9906550-A2.  
XX  
XX 11-FEB-1999.  
XX  
XX 31-JUL-1998; 98WO-IB01232.  
XX  
XX 01-AUG-1997; 97US-0905144.  
XX  
XX (GEST ) GENSET.  
XX  
XX Duclet A, Dumas Milne Edwards J, Lacroix B;  
XX  
XX WPI: 1999-153780/13.  
XX  
XX N-PSDB: AAX40642.  
XX  
XX New isolated prostate-derived nucleic acids - used to develop  
XX products which may have cytokine, immune regulatory, haematopoiesis  
XX regulating, anti-inflammatory or tumour inhibition activity  
XX  
XX Claim 34: Page 629; 675pp; English.  
XX  
XX AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for  
XX human secreted proteins expressed in prostate, and encode the proteins  
XX given in AAV11716 to AAV11993 respectively. The proteins given represent  
XX the signal peptide and an N-terminal fragment of a secreted protein. The  
XX nucleic acid sequences can be used for producing secreted human gene  
XX products. They can also be used to develop products for diagnosis and  
XX therapy. The proteins obtained may have cytokine activity, cell  
XX proliferation and differentiation activity, haematopoiesis regulating  
XX activity, tissue growth regulating activity, reproductive hormone  
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and  
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory  
XX activity, tumour inhibition activity or other activities. The products  
XX can be used in forensic, gene therapy and chromosome mapping procedures.

CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptides can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
XX polypeptide into a membrane, or importing a polypeptide into a cell.  
XX  
SQ Sequence 124 AA:  
  
Query Match 44.1%; Score 45; DB 20; Length 124;  
Best Local Similarity 52.6%; Pred. No. 11;  
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
OY 2 YKRLRYGKRLRYAOEE 20  
|||  
Db 87 ykrlkaevkqskkylmve 105  
  
RESULT 23  
AAV34690  
ID AAV34690 standard; Protein: 214 AA.  
XX  
XX AAV34690;  
XX  
XX 13-SEP-1999 (first entry)  
XX  
XX Chlamydia pneumoniae transmembrane protein sequence.  
XX  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
XX vaccine; neutralising epitope.  
XX  
XX Chlamydia pneumoniae.  
XX  
XX WO9927105-A2.  
XX  
XX 03-JUN-1999.  
XX  
XX 20-NOV-1998; 98WO-IB01890.  
XX  
XX 04-NOV-1998; 98US-0107078.  
XX  
XX 21-NOV-1997; 97FR-0014673.  
XX  
XX (GEST ) GENSET.  
XX  
XX Griffais R;  
XX  
XX WPI: 1999-357842/30.  
XX  
XX Genome sequence of Chlamydia pneumoniae  
XX  
XX Page 688-689; Disclosure: 1912pp; English.  
XX  
XX AAV34584-Y35879 represent the proteins encoded by all the open reading  
XX frames in the complete genome (see AAV31990) of Chlamydia pneumoniae.  
XX C. pneumoniae causes respiratory disease such as pneumonia and  
XX bronchitis and is thought to be a contributing factor in heart  
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
XX nodosum or pharyngitis. The polypeptides encoded by the open reading  
XX frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in  
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
XX nucleotide sequences can also be used as immunogenic compositions,  
XX especially where the vector directs the expression of a neutralising  
XX epitope of C. pneumoniae.  
XX  
SQ Sequence 214 AA:  
  
Query Match 44.1%; Score 45; DB 20; Length 214;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
OY 3 KRLRYGKRLRYAOEE 20  
|||  
KRLRYGKRLRYAOEE 20

PF 14-JUL-2000; 2000MO-US19267.  
XX  
PR 15-JUL-1999; 99US-0144412.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
XX detecting anti-hepatitis A virus and as vaccines -  
XX  
PS Claim 13; Page 94; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
XX  
SQ Sequence 21 AA:  
  
Query Match 44.18; Score 45; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 12 QRLRYAOEE 20  
Db 1 qrlkyagee 9  
  
RESULT 20  
AAW42930  
ID AAW42930 standard; peptide; 25 AA.  
AC  
XX AAW42930;  
XX  
DT 28-APR-1998 (first entry)  
XX  
XX Immunogenic Hepatitis A virus peptide YK-1665.  
DE  
XX Immunogenic Hepatitis A virus peptide YK-1665.  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW Immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN W09740147-A1.  
XX  
PD 30-OCN-1997.  
XX  
PF 18-APR-1997; 97MO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.

XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
XX Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. Compositions  
CC containing the peptides can be used to induce an immune response to HAV  
CC in a mammal. The peptides can also be used to detect the presence of  
CC antibodies against HAV in mammalian serum. The peptides can also be used  
CC to make an antibody against HAV by administering the peptide to a  
CC mammal.  
XX  
SQ Sequence 25 AA:  
  
Query Match 44.18; Score 45; DB 18; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 12 QRLRYAOEE 20  
Db 1 qrlkyagee 9  
  
RESULT 21  
AAB69447  
ID AAB69447 standard; Peptide; 26 AA.  
AC  
XX AAB69447;  
XX  
DT 20-APR-2001 (first entry)  
XX  
XX Synthetic HAV P2A peptide, SEQ ID NO: 47.  
DE  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; vaccine; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
PN W0200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000MO-US19267.  
XX  
PR 15-JUL-1999; 99US-0144412.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
XX  
DR WPI; 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines -  
XX  
PS Claim 13; Page 98; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a

Db 8 pykelrlevyqkr 20

RESULT 17  
AAB69439  
ID AAB69439 standard; Peptide: 21 AA.  
XX  
AC AAB69439;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Synthetic HAV P2A peptide, SEQ ID NO: 39.  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
FN WO200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PE 14-JUL-2000; 2000WO-US19267.  
XX  
PR 15-JUL-1999; 99US-0144412.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI: 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines -  
XX  
PS Claim 13; Page 93; 130pp; English.  
XX  
CC The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
XX  
SQ Sequence 21 AA:  
  
Query Match 65.7%; Score 67; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.00043;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGRKOR 13  
XXXXXXXXXXXXXXXXXXXX  
Db 8 pykelrlevyqkr 20

RESULT 18  
AAM42924  
ID AAM42924 standard; peptide: 20 AA.  
XX  
AC AAM42924;  
XX  
PD 25-JAN-2001.

DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1317.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PE 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI: 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
CC Peptides AAM42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 810-829, and has a reactivity of 83.3% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal.  
XX  
SQ Sequence 20 AA:  
  
Query Match 44.1%; Score 45; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 QRLKXQDEE 20  
XXXXXXXXXXXX  
Db 1 qrlkyaqee 9

RESULT 19  
AAB69441  
ID AAB69441 standard; Peptide: 21 AA.  
XX  
AC AAB69441;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Synthetic HAV P2A peptide, SEQ ID NO: 41.  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
FN WO200105824-A2.  
XX  
PD 25-JAN-2001.



CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.

XX Sequence 2227 AA;

Query Match 100.0%; Score 102; DB 21; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLKYAOEE 20

Db 799 pykeirlevgkgrlkyagee 818

RESULT 13

AAB18609 AAB18609 standard; Protein; 2227 AA.

XX AAB18609;

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
KM HAV 4380.

OS Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PE 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93MO-US08610.

PR 10-MAR-1995; 95US-0397232.

PA (USSR ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI: 2000-586464/55.

DR N-PSDB; AAA75478.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX Disclosure; Columns 93-104; 72pp; English.

XX The present sequence is derived from a live attenuated hepatitis A

CC virus (HAV) of the invention, designated HAV 4380. The sequence is

CC produced by modifying wild type HAV strain HM-174. The HAV of the

CC invention are adapted to growth in the human fibroblast-like cell

CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain

CC appropriate attenuation. It is useful as a live vaccine for prophylaxis

CC of hepatitis A in humans and other primates.

XX Sequence 2227 AA;

QY 1 PYKEIRLEVGKORLKYAOEE 20

Db 799 pykeirlevgkgrlkyagee 818

RESULT 14

AAP60066 AAP60066 standard; Protein; 2227 AA.

XX AAP60066;

DT 26-JUN-1991 (first entry)

DE Sequence of viral I434 polypeptide encoded by the complete

DE nucleotide sequence of the HAV genome.

XX Diagnosis; vaccine; passive immunotherapy.

OS Hepatitis A virus.

FT Key Location/Qualifiers

FT 1..245 /label= P1.1A

FT 246..491 /label= 1B

FT 492..836 /label= 1C

FT 837..980 /label= P2.2A

FT 981..1076 /label= 2B

FT 1077..1422 /label= 2C

FT 1423..1484 /label= P3.3A

FT 1485..1507 /label= 3B

FT 1508..1678 /label= 3C

FT 1679..2227 /label= 3D

XX EPI99480-A.

XX 29-OCT-1986.

PE 03-APR-1986; 86EP-0302465.

PR 03-APR-1985; 85US-0719329.

PA (CHIR-) CHIRON CORP.

PI Dina D, Potter SJ, Vannest GA, Caput D;

DR WPI: 1986-286213/44.

DR N-PSDB; AAP60080.

XX Hepatitis A virus nucleotide sequence and polypeptide - and use

PT in prodn. of vaccines and diagnostic probes

XX Claim 5; Fig 1; 18pp; English.

XX AAP60080 and oligonucleotide fragments are useful in detection of

CC hepatitis A virus; transformed hosts may be used for expression of

CC polypeptides and fragments useful in vaccines without risk of

CC infection by the virus or in prodn. of particles which are capable

CC of inducing immunocompetent B cells for passive immunotherapy. Pref.

CC epitope is derived from AAs 445-657 or 792-848 of the HAV

XX polypeptide sequence (AAP60066).

XX Sequence 2227 AA;

QY 1 PYKEIRLEVGKORLKYAOEE 20

Query Match 97.1%; Score 99; DB 7; Length 2227;



XX W09740166-A2.  
 XX 30-OCT-1997.  
 XX 18-APR-1997; 97WO-US06506.  
 XX 19-APR-1996; 96US-0015642.  
 XX (USSH ) US SEC DEPT HEALTH.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Emerson SU, Purcell RH, Raychaudhuri G;  
 XX WPI: 1997-535850/49.  
 XX N-PSDB: AAT93023.  
 XX Human attenuated HAV genome containing simlan HAV 2C gene - useful  
 XX as vaccines against HAV infection  
 XX Disclosure: Fig 13A-D; 66pp; English.  
 XX  
 CC This protein sequence is encoded by the human hepatitis A virus  
 CC (HAV) HM-175 wild-type genome (see AAT93023). Attenuated strain  
 CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
 CC kidney cells. A claimed DNA construct (1) comprises a genome of  
 CC HAV, where the genome is a human attenuated HAV genome in which a  
 CC region of the 2C gene has been replaced by a corresponding region  
 CC from a 2C gene of a simlan AGM-27 HAV genome (see AAT93024). The  
 CC region of the 2C gene from AGM-27 contained in the construct  
 CC preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
 CC transcript of (1); (2) a cell transfected with (1) or the RNA  
 CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
 CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
 CC its RNA transcript, can be used as a vaccine for preventing HAV in  
 CC a mammal. (1) or the RNA transcript can also be used to stimulate  
 CC the production of protective antibodies in the mammal.  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 102; DB 18; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKEIRLEVGKORLKYAOEE 20  
 DB 799 PYKEIRLEVGKGRILKYAGEE 818  
 RESULT 11  
 AAB18607  
 ID AAB18607 standard; Protein: 2227 AA.  
 AC AAB18607;  
 XX 15-JAN-2001 (first entry)  
 DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 XX Hepatitis A virus.  
 XX US6113912-A.  
 XX 05-SEP-2000.  
 XX 07-JUN-1995; 95US-0475886.  
 XX 18-SEP-1992; 92US-0947338.  
 XX 17-SEP-1993; 93WO-US08610.

PR 10-MAR-1995; 95US-0397232.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX WPI: 2000-586464/55.  
 XX N-PSDB: AAA75476.  
 XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 XX line useful as vaccine for protecting humans against hepatitis A virus  
 XX infection, has modified genome compared to wild type  
 XX Disclosure: Fig 6A-K; 72pp; English.  
 XX  
 CC The present sequence is derived from a wild type hepatitis A virus  
 CC (HAV) strain HM-174. The sequence is modified to produce HAV which  
 CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
 CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
 CC attenuation. It is useful as a live vaccine for prophylaxis of  
 CC hepatitis A in humans and other primates.  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 102; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKEIRLEVGKORLKYAOEE 20  
 DB 799 PYKEIRLEVGKGRILKYAGEE 818  
 RESULT 12  
 AAB18608  
 ID AAB18608 standard; Protein: 2227 AA.  
 AC AAB18608;  
 XX 15-JAN-2001 (first entry)  
 DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
 XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 XX P-35 virus.  
 XX Hepatitis A virus.  
 XX US6113912-A.  
 XX 05-SEP-2000.  
 XX 07-JUN-1995; 95US-0475886.  
 XX 18-SEP-1992; 92US-0947338.  
 XX 17-SEP-1993; 93WO-US08610.  
 XX 10-MAR-1995; 95US-0397232.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX WPI: 2000-586464/55.  
 XX N-PSDB: AAA75477.  
 XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 XX line useful as vaccine for protecting humans against hepatitis A virus  
 XX infection, has modified genome compared to wild type  
 XX Disclosure: Columns 67-78; 72pp; English.  
 XX The present sequence is derived from passage 35 of a wild type

in marmosets). The RNA was used to prepare ds cDNA clones by standard methods. Clones containing inserts which hybridised to RNA from HAV-infected African Green Monkey Kidney cells were selected for further analysis. A 7.4kb restriction map (about 99% of the HAV genome) was constructed from 5 overlapping inserts. The sequence of the first 3.3kb (approx.) from the 5'-terminus was determined. An amino acid sequence was deduced from the entire clone and an open reading frame was identified starting at position 238. A comparison of the predicted HAV amino acid sequences with the known capsid protein sequences of other picornaviruses (poliovirus, foot and mouth disease virus and encephalomyelitis virus) revealed areas of local homology.

(Note: Revised entry submitted to correct the patent number format of US Government-owned NTIS applications to prevent clashes with ongoing US granted patent numbers. For further information please visit the Derwent web site at [www.derwent.com/dwpi/updates/ntis\\_us.html](http://www.derwent.com/dwpi/updates/ntis_us.html).)

CC US4894228-A.  
XX 16-JAN-1990.  
XX 12-JUL-1988; 88US-0217824.

Query Match 100.0%; Score 102; DB 14; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 5.8e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKRLREVGKORLKYAOEE 20  
DB 1036 PYKRLREVGKQRIKYAGEE 1055

RESULT 9  
AA05697 standard; protein: 2227 AA.

XX AAR05697;  
XX 15-AUG-1990 (first entry)  
XX Attenuated hepatitis A virus.  
XX Hepatitis A virus; vaccine; attenuated.  
XX Hepatitis A virus, strain HM-175.  
XX OS  
XX FH Key Location/Qualifiers  
FT Region 1..23  
FT /label=VP4 = 1A  
FT 24..245  
FT /label=VP2 = 1B  
FT 246..491  
FT /label=VP3 = 1C  
FT 492..791  
FT /label=VP1 = 1D  
FT 792..980  
FT /label=2A  
FT 981..1087  
FT /label=2B  
FT 1088..1422  
FT /label=2C  
FT 1423..1496  
FT /label=3A  
FT 1497..1519  
FT /label=3B = VPg  
FT 1520..1738  
FT /label=3C  
FT 1739..2227  
FT /label=3D  
XX  
XX PN US4894228-A.  
XX 16-JAN-1990.  
XX 12-JUL-1988; 88US-0217824.

PR 12-JUL-1988; 88US-0217824.  
PR 12-JUL-1988; 88US-0652967.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN.  
XX  
XX Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;  
PI Daemer RJ, Gust ID;  
DR WPI; 1990-075557/10.  
DR N-PSDB; AAO03512.  
XX  
XX Vaccine against hepatitis A virus infection - comprises novel  
XX attenuated hepatitis A virus strain.  
XX  
XX Claim 1; Fig 1; 18pp; English.

CC The attenuated HAV is useful for inducing protective immunity against  
CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
CC several nucleotide changes distributed throughout the genome, is  
CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
CC suitable for use as an HAV vaccine. It is noted that not all the changes  
CC are necessary for attenuation and use as a vaccine.

SQ Sequence 2227 AA;

Query Match 100.0%; Score 102; DB 11; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKRLREVGKORLKYAOEE 20  
DB 799 PYKRLREVGKQRIKYAGEE 818

RESULT 10  
AAW34074 standard; Protein: 2227 AA.

XX AAW34074;  
XX 27-APR-1998 (first entry)  
XX Hepatitis A virus HM-175 protein sequence.  
XX DE Hepatitis A virus; vaccine.  
XX KW HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;  
XX infection; vaccine.  
XX OS Hepatitis A virus HM-175.  
XX FH Key Location/Qualifiers  
FT Protein 1..23  
FT /label= VP4  
FT 24..245  
FT /label= VP2  
FT 246..491  
FT /label= VP3  
FT 492..791  
FT /label= VP1  
FT 792..980  
FT /label= 2A  
FT 981..1087  
FT /label= 2B  
FT 1088..1422  
FT /label= 2C  
FT 1423..1496  
FT /label= 3A  
FT 1497..1519  
FT /label= 3B  
FT 1520..1738  
FT /label= 3C  
FT 1739..2227  
FT /label= 3D  
FT Protein

```

PN      EPI38704-A.
XX
XX      24-APR-1985.
XX
XX      09-OCT-1984;      84EP-0402025.
XX
XX      02-MAR-1984;      84US-0585942.
XX      14-OCT-1983;      83US-0541836.
XX
XX      (MERI ) MERCK & CO INC.
XX
XX      Hughes JV, Scolnick EM, Tomassini JE;
XX
XX      WPI; 1985-100818/17.
XX      DR      N-PSDB; AAN50274.
XX
XX      New hepatitis A virus surface protein - useful for binding to
XX      neutralising antibodies to the virus
XX
XX      Disclosure; Page 17-23; 49pp; English.
XX
XX      VPI is isolated by solubilisation of the intact virus in an aq.
XX      anionic surfactant and a reducing agent. The viral proteins are sepd.
XX      and the protein of molecular wt. 33000 daltons is sepd.
XX
XX      Sequence      993 AA;
XX
XX      Query Match      100.0%; Score 102; DB 6; Length 993;
XX      Best Local Similarity 100.0%; Pred. No. 5,3e-08;
XX      Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY      1 PYKEIRLEVGKORLKYAEE 20
XX      ||||||||||||||||
DB      935 PYKEIRLEVgkqrlkyagee 954

RESULT 7
AAW95559
ID      AAW95559 standard; Protein; 1077 AA.
XX
XX      AAW95559;
XX
XX      28-APR-1999 (first entry)
XX
XX      A partial hepatitis A virus (HAV) protein.
XX
XX      Hepatitis A virus protein; HAV; P2 region;
XX      cell-culture-adapted HAV strain; Infection; accelerated growth.
XX
XX      Hepatitis A virus.
XX
XX      US5849562-A.
XX
XX      15-DEC-1998.
XX
XX      06-JUN-1995;      95US-0468926.
XX
XX      06-NOV-1991;      91US-0788262.
XX      30-SEP-1983;      83US-0537911.
XX      27-SEP-1984;      84US-0654942.
XX      06-OCT-1984;      88US-0256135.
XX      06-JUN-1995;      95US-0468926.
XX
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX      Emerson SU, Purcell RH;
XX
XX      WPI; 1999-094412/08.
XX      DR      N-PSDB; AAX01006.
XX
XX      Chimeric hepatitis A virus strains - with P2 region from
XX      cell-culture-adapted strain in wild-type genome
XX

```

```

XX      Disclosure; Fig 7A-L; 36pp; English.
XX
XX      The present sequence represents a partial hepatitis A virus (HAV)
XX      protein. The specification describes a DNA construct consisting of
XX      of a wild-type HAV genome in which the P2 region is replaced by the
XX      P2 region from a cell-culture-adapted HAV strain. The construct is
XX      used to demonstrate that mutations in the P2 region of a
XX      cell-culture-adapted HAV strain are sufficient for establishment of
XX      infection and accelerated growth in cell culture.
XX
XX      Sequence      1077 AA;
XX
XX      Query Match      100.0%; Score 102; DB 20; Length 1077;
XX      Best Local Similarity 100.0%; Pred. No. 5,8e-08;
XX      Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY      1 PYKEIRLEVGKORLKYAEE 20
XX      ||||||||||||||||
DB      1022 PYKEIRLEVgkqrlkyagee 1041

RESULT 8
AAR32426
ID      AAR32426 standard; Protein; 1091 AA.
XX
XX      AAR32426;
XX
XX      17-DEC-2001 (updated)
XX      DT      10-JUN-1993 (first entry)
XX
XX      Translated from 5' region of Hepatitis A Virus genomic clone.
XX
XX      HAV HM-175; chronic liver disease; picornavirus.
XX
XX      Hepatitis A Virus.
XX
XX      Location/Qualifiers
XX      FH      238..1091
XX      FT      Region
XX      FT      /label= ORF
XX      FT      /note= "second putative initiation codon at
XX      position 240"
XX      FT      1..711
XX      FT      /note= "X's correspond to nonsense codons,
XX      i.e. this region is not an ORF"
XX
XX      USN7788262-N.
XX
XX      15-DEC-1992.
XX
XX      30-SEP-1983;      83US-0536911.
XX
XX      27-SEP-1984;      84US-0654942.
XX      06-OCT-1988;      88US-0256135.
XX      30-SEP-1983;      83US-0536911.
XX      06-NOV-1991;      91US-0788262.
XX
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX      Baltimore D, Feinstone SM;
XX      PI      Purcell RH, Racanelli VR, Ticehurst JR;
XX
XX      WPI; 1993-067429/08.
XX      DR      N-PSDB; AAO36934.
XX
XX      Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.
XX      of antigen and antibodies
XX
XX      Disclosure; Fig 7; 65pp; English.
XX
XX      HAV virion RNA was extracted from the livers of marmosets which had
XX      been inoculated with HAV (the HAV had previously been passaged twice
XX

```

Query Match 100.0%; Score 102; DB 6; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEIRLEVGKORLKYAEE 20  
 |||  
 DB 308 PYKEIRLEVgKqrlkyagee 327

RESULT 4  
 AAP50287  
 ID AAP50287 standard; Protein; 854 AA.  
 XX  
 AC AAP50287;  
 XX  
 DT 30-NOV-1991 (first entry)  
 XX  
 DE Sequence encoded by hepatitis A virus (HAV) CDNA from near the  
 DE genome 5' terminus to the end of the area corresponding to the  
 DE capsid protein region of poliovirus RNA.  
 XX  
 KM Hepatitis A virus assay; antigen; antibody.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN W08501517-A.  
 XX  
 PD 11-APR-1985.  
 XX  
 PF 27-SEP-1984; 84WO-0501552.  
 XX  
 PR 30-SEP-1983; 83US-0537911.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECH.  
 XX  
 PI Titechurst JR, Baltimore D, Reinstone SM, Purcell RH;  
 PI Racanelli VR;  
 XX  
 DR WPI; 1985-098846/16.  
 DR N-PSDB; AAN50330.  
 XX  
 PT New hepatitis A virus CDNA - useful in assays for the virus and  
 PT for prodn. of the viral antigen and antibodies to it  
 XX  
 PS Example; Fig 7; 60pp; English.  
 XX  
 CC The inventors claim HAV CDNA and a method for producing it, whereby  
 CC large amts. can be obtd. economically. The CDNA is useful in the  
 CC assay for detection of HAV quickly and easily and with high  
 CC sensitivity and specificity. The HAV CDNA is also used in the prodn.  
 CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.  
 XX  
 SO Sequence 854 AA.

Query Match 100.0%; Score 102; DB 6; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEIRLEVGKORLKYAEE 20  
 |||  
 DB 799 PYKEIRLEVgKqrlkyagee 818

RESULT 5  
 AAP50116  
 ID AAP50116 standard; Protein; 993 AA.  
 XX  
 AC AAP50116;  
 XX  
 DT 30-SEP-1991 (first entry)  
 XX

DE Sequence of Hepatitis A virus (HAV) immunogenic peptides  
 DE VP-1, VP-2, VP-3 and VP-4.  
 XX  
 KM Antigenic protein; immunogen; vaccine.  
 XX  
 OS Hepatitis A virus (strain CR326).  
 XX  
 PN EP154587-A.  
 XX  
 PD 11-SEP-1985.  
 XX  
 PF 27-FEB-1985; 85EP-0400369.  
 XX  
 PR 02-MAR-1984; 84US-0585818.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Linemeyer DL, Menke JG, Reuben RG, Mitra SW;  
 XX  
 DR WPI; 1985-224964/37.  
 DR N-PSDB; AAN50139.  
 XX

PT New nucleotide sequences coding for hepatitis A virus antigens -  
 PT useful for eliciting normal immune response and in vaccines for  
 PT protecting against the virus  
 XX  
 PS Example; Page 11-17; 32pp; English.  
 XX  
 CC Within the sequence in AAN50139 is encoded the information necessary  
 CC to make the antigenic proteins of HAV. The sequences encoding for  
 CC the structural proteins begin at base 403. The key sub-unit  
 CC sequences within VP-1, designated Sequences I, II, III, IV, and V,  
 CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other  
 CC nucleotide sequences which are valuable as encoding antigenic  
 CC proteins are the sequences from base 1749 to base 2722; from base  
 CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from  
 CC base 1749 to base 2722 is esp. valuable as a vector for producing  
 CC antigen protein. Sequences II-V are claimed. X in AAP50116 denotes the  
 CC translation of a stop codon.  
 XX  
 SO Sequence 993 AA.

Query Match 100.0%; Score 102; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKEIRLEVGKORLKYAEE 20  
 |||  
 DB 935 PYKEIRLEVgKqrlkyagee 954

RESULT 6  
 AAP50231  
 ID AAP50231 standard; Protein; 993 AA.  
 XX  
 AC AAP50231;  
 XX  
 DT 28-NOV-1991 (first entry)  
 XX  
 DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
 DE including surface protein (VP-1).  
 XX  
 KM Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KM diagnostic assay.  
 XX  
 OS Hepatitis A virus.  
 XX  
 FH Protein  
 FT 628..993  
 FT /note="claimed; X denotes translated stop codons  
 FT and unspecified triplets"

XX	18-APR-1997;	97WO-US06891.
PF		
XX	19-APR-1996;	96US-0015644.
PR		
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
PA		
XX		
PI	Fields HA, Khudyakov YE;	
DR	WPI: 1997-535831/49.	
XX		
PT	Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an	
PT	immune response to HAV in a mammal or to detect the presence of	
PT	antibodies against HAV in a mammal	
XX		
PS	Claim 18; page 112; 140pp; English.	
XX		
CC	Peptides AAW42922-30 are immunogenic peptides corresponding to	
CC	immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are	
CC	substantially similar to a portion of the amino acid sequence of the P2A	
CC	protein of HAV corresponding to amino acids 797-980. The present peptide	
CC	is derived from amino acids 799-818, and has a reactivity of 41.7% with	
CC	acute sera. Compositions containing the peptides can be used to	
CC	immune response to HAV in a mammal. The peptides can also be used to	
CC	detect the presence of antibodies against HAV in mammalian serum. The	
CC	peptides can also be used to make an antibody against HAV by	
CC	administering the peptide to a mammal.	
XX		
SO	Sequence 20 AA;	
XX		
XX	Query Match	100.0%; Score 102; DB 18; Length 20;
XX	Best Local Similarity	100.0%; Pred. No. 8.4e-10;
XX	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 PYKEIRLEVGKQRLKYAOGE 20	
DB	1 pykeirlevgkqrlkyagee 20	
XX		
XX	RESULT 2	
XX	AAB69440	
ID	AAB69440 standard; Peptide: 21 AA.	
XX		
AC	AAB69440;	
XX		
DT	20-APR-2001 (first entry)	
XX		
DE	Synthetic HAV P2A peptide, SEQ ID NO: 40.	
XX		
XX	Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;	
XX	antigen; major structural capsid polypeptide; HAV antibody detection.	
XX		
OS	Hepatitis A virus.	
OS	Synthetic.	
XX		
PN	WO200105824-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	14-JUL-2000; 2000WO-US19267.	
XX		
PR	15-JUL-1999; 99US-0144412.	
XX		
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Fields HA, Khudyakov YE;	
XX		
DR	WPI: 2001-112681/12.	
XX		
PT	Synthetic peptides used as antigen sources for enzyme immunoassays	
PT	detecting anti-hepatitis A virus and as vaccines -	

PS Claim 13, Page 93; 130pp; English.

CC The present sequence is one of a number of synthetic peptides which are  
XX immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
XX

SQ Sequence 21 AA:

QY 1 PYKEIRLEVGKQRILKYADEE 20  
|||  
Db 1 pykelrlvvgkqrilkyaqee 20

RESULT 3  
AAP50230 ID AAP50230 standard; Protein: 366 AA.  
AC AAP50230:  
XX 28-NOV-1991 (first entry)  
DT XX  
DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).  
XX  
KW Hepatitis A virus vaccine; Immunisation; monoclonal antibody;  
KM diagnostic assay.  
XX  
OS Hepatitis A virus.  
XX  
EP138704-A.  
FN  
XX 24-APR-1985.  
PD  
XX 09-OCT-1984; 84EP-0402025.  
PF  
XX 02-MAR-1984; 84US-0585942.  
PR  
XX 14-OCT-1983; 83US-0541836.  
RR  
XX (MERI ) MERCK & CO INC.  
PA  
XX Hughes JV, Scolnick EM, Tomassini JE;  
PI  
DR WPI; 1985-100818/17.  
N-PSDB; AAN50274.  
DR  
XX New hepatitis A virus surface protein - useful for binding to  
PT neutralising antibodies to the virus  
PT  
XX Claim 21; Page 46-48; 49pp; English.  
PS  
XX VPI is isolated by solubilisation of the intact virus in an aq.  
CC antionic surfactant and a reducing agent. The viral proteins are sepd.  
CC and the protein of molecular wt. 33000 daltons is sepd.  
XX  
SQ Sequence 366 AA:

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:01:40 ; Search time 209.1 Seconds  
(without alignments)  
10.624 Million cell updates/sec

```
Title: US-09-171-432A-40
Perfect score: 102
Sequence: 1 PYKELRLVGVKQRLKYAQEE 20
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 60 summaries
```

Database : A\_Geneseq\_032802.\*

1:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDS5/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	102	100.0	20	18	AAW42923	Immunogenic Hepati
2	102	100.0	21	22	AAB69440	Synthetic HAV p2A
3	102	100.0	366	6	AA50330	Sequence of hepati
4	102	100.0	854	6	AA50287	Sequence encoded b
5	102	100.0	993	6	AA50116	Sequence of Hepati
6	102	100.0	993	6	AA50231	Sequence encoded b
7	102	100.0	1077	20	AAW95559	A partial hepatiti
8	102	100.0	1091	14	AAW32426	Translated from 5'
9	102	100.0	2227	11	AAW05697	Attenuated hepatiti
10	102	100.0	2237	18	AAW34074	Hepatitis A virus
11	102	100.0	2227	21	AAW18607	Amino acid sequenc

12	102	100.0	2227	21	AA818609
13	102	100.0	2227	21	AA818609
14	99	97.1	2227	7	AAPE0066
15	97	95.2	839	18	AAAI252229
16	94	92.7	20	12	AAAI292222
17	67	65.7	21	22	AA869439
18	45	44.1	20	18	AA8A29242A
19	45	44.1	21	22	AA869441
20	45	44.1	25	18	AA8A293038
21	45	44.1	26	22	AA869447
22	45	44.1	124	20	AA81192020
23	45	44.1	214	20	AA83469020
24	44	43.1	211	22	AA85293737
25	44	43.1	211	22	AA852971
26	44	43.1	1139	22	AA865314
27	44	43.1	1503	22	AA86509444
28	43	42.2	120	22	AA852798
29	43	42.2	178	22	AA80366666
30	42	41.2	182	22	AA856446
31	42	41.2	207	22	AA852857
32	42	41.2	394	22	AA8G103404
33	42	41.2	411	21	AA8G397939
34	42	41.2	432	22	AA8E110655
35	42	41.2	452	21	AA8G397922
36	42	41.2	900	22	AA8G386488
37	42	41.2	1507	22	AA8G303433
38	41	40.2	46	22	AA8B284040
39	41	40.2	46	22	AA8B335884
40	41	40.2	46	22	AA8B19041
41	41	40.2	46	22	AA8A270451
42	41	40.2	46	22	AA8A234949
43	41	40.2	46	22	AA8A675359
44	41	40.2	46	22	AA8A462422
45	41	40.2	46	22	AA867585
46	41	40.2	68	18	AA809417
47	41	40.2	68	18	AA809417
48	41	40.2	68	18	AA8A1908
49	41	40.2	68	20	AA8A179866
50	41	40.2	68	20	AA8A406666
51	41	40.2	79	21	AA8A343332
52	41	40.2	88	21	AA800165
53	41	40.2	147	17	AA888419
54	41	40.2	147	21	AA86290053
55	41	40.2	161	21	AA8B32201
56	41	40.2	187	21	AA8B5341
57	41	40.2	215	22	AA8G380747
58	41	40.2	397	22	AA8E048866
59	41	40.2	404	22	AA8E4873
60	41	40.2	420	22	AA8B63362

## ALIGNMENTS

RESULT	1
AAW42923	
ID	AAW42923 standard; peptide; 20 AA.
XX	
XX	AAW42923;
AC	
XX	
DT	28-APR-1998 (first entry)
XX	
DE	
XX	Immunogenic Hepatitis A virus peptide YK-1316.
XX	
KW	Immunogenic peptide; Immunogenic epitope; P2A protein.
KW	Immune response; antibody.
XX	
OS	Synthetic.
OS	Hepatitis A virus.
XX	
XX	W09740147-A1.
FN	
XX	
PD	30-OCT-1997.

FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MM; 8438C51846AEFAA4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGR 48

Search completed: June 16, 2002, 00:08:51  
Job time: 789 sec

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGR 48

## RESULT 22

039873 ID 039873 PRELIMINARY; PRT; 56 AA.  
AC 039873;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-2333;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
RT 1982-1996."  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL; U68698; AAB53594.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGR 48

## RESULT 23

039874 ID 039874 PRELIMINARY; PRT; 56 AA.  
AC 039874;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-314274;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
RT 1982-1996."  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL; U68699; AAB53595.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGR 48

## RESULT 24

067818 ID 067818 PRELIMINARY; PRT; 56 AA.  
AC 067818;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISR-05, AFULA DISTRICT, ISRAEL, 1993;  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
RT Israel, based on their VP1/2A nucleotide sequence."  
RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; Z77243; CAB01036.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 1  
FT CHAIN 29  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGR 48

## RESULT 25

067819 ID 067819 PRELIMINARY; PRT; 56 AA.  
AC 067819;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISR-10, AFULA DISTRICT, ISRAEL, 1993);  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
RT Israel, based on their VP1/2A nucleotide sequence."  
RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; Z77244; CAB01037.1; -;  
DR InterPro: IPR000886; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 1  
FT CHAIN 29  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;





Db 29 SHIECRKPYKELRLEVGR 48

RESULT 14

039866 ID 039866 PRELIMINARY; PRT; 56 AA.  
 AC 039866; 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=96001190;  
 RX MEDLINE=97247817; PubMed=9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL; U68690; AAB53586.1; -;  
 DR InterPro; IPR000886; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6614 MM; 8438C51846AEFA4A CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECRKPYKELRLEVGR 20  
 Db 29 SHIECRKPYKELRLEVGR 48

RESULT 15

096829 ID 096829 PRELIMINARY; PRT; 56 AA.  
 AC 096829; 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=406909;  
 RX MEDLINE=97247817; PubMed=9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL; U68691; AAB53587.1; -;  
 DR InterPro; IPR000886; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6623 MM; A917151846AEFA47 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECRKPYKELRLEVGR 20

Db 29 SHIECRKPYKELRLEVGR 48

RESULT 16

039867 ID 039867 PRELIMINARY; PRT; 56 AA.  
 AC 039867; 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JVR;  
 RX MEDLINE=97247817; PubMed=9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL; U68692; AAB53588.1; -;  
 DR InterPro; IPR000886; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6614 MM; 8438C51846AEFA4A CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECRKPYKELRLEVGR 20  
 Db 29 SHIECRKPYKELRLEVGR 48

RESULT 17

039868 ID 039868 PRELIMINARY; PRT; 56 AA.  
 AC 039868; 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=503712;  
 RX MEDLINE=97247817; PubMed=9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL; U68693; AAB53589.1; -;  
 DR InterPro; IPR000886; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SO SEQUENCE 56 AA; 6614 MM; 8438C51846AEFA4A CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
RESULT 10
099974 ID 099974 PRELIMINARY; PRT; 55 AA.
AC 099974;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHILE-3;
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and
co-circulation during epidemic outbreaks.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ306384; CAC29233.1; -.
DR InterPro: IPR000886; ER_TARGET.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match 100.0%; Score 107; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKOR 20
Db 28 SHIECRKPKYKELRLEVGKOR 47
|||||
```

```
RESULT 12
099972 ID 099972 PRELIMINARY; PRT; 55 AA.
AC 099972;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UR03;
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and
co-circulation during epidemic outbreaks.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ306386; CAC29235.1; -.
DR InterPro: IPR000886; ER_TARGET.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match 100.0%; Score 107; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKOR 20
Db 28 SHIECRKPKYKELRLEVGKOR 47
|||||
```

```
RESULT 13
039865 ID 039865 PRELIMINARY; PRT; 56 AA.
AC 039865;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=412991;
RX MEDLINE=97247817; PubMed=9093940;
RA Taylor M.B.;
RT "Molecular epidemiology of South African strains of hepatitis A virus:
1982-1996.";
RL J. Med. Virol. 51:273-279(1997).
DR EMBL: U66889; AAB53585.1; -.
DR InterPro: IPR000886; ER_TARGET.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKOR 20
|||||
```

QY 1 SHIECKRPPKELRLEVGR 20  
DB 29 SHIECKRPPKELRLEVGR 48

RESULT 6  
Q98VY2 PRELIMINARY: PRT: 53 AA.  
AC Q98VY2; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URUGUAY4;  
RA Costa-Matoli M., Ferre V., Monpocho S., Garcia L., Collina R.,  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ309234; CAC37078.1; -.  
FT NON\_TER 1 53  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKRPPKELRLEVGR 20  
DB 29 SHIECKRPPKELRLEVGR 48

RESULT 7  
Q98VY1 PRELIMINARY: PRT: 53 AA.  
AC Q98VY1; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHILE16;  
RA Costa-Matoli M., Ferre V., Monpocho S., Garcia L., Collina R.,  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ309233; CAC37079.1; -.  
FT NON\_TER 1 53  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKRPPKELRLEVGR 20  
DB 29 SHIECKRPPKELRLEVGR 48

RESULT 8  
Q98VY0 PRELIMINARY: PRT: 53 AA.  
AC Q98VY0; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URUGUAY4;  
RA Costa-Matoli M., Ferre V., Monpocho S., Garcia L., Collina R.,  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ309234; CAC37080.1; -.  
FT NON\_TER 1 53  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKRPPKELRLEVGR 20  
DB 29 SHIECKRPPKELRLEVGR 48

RESULT 9  
Q99VU8 PRELIMINARY: PRT: 55 AA.  
AC Q99VU8; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARG-6;  
RA Costa-Matoli M., Ferre V., Monpocho S., Garcia L., Collina R.,  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ306370; CAC29219.1; -.  
DR InterPro: IPR000886; ER-target.  
DR PROSITE: PS00014; ER-TARGET; UNKNOWN\_1.  
FT NON\_TER 1 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4C8 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKRPPKELRLEVGR 20  
DB 28 SHIECKRPPKELRLEVGR 47

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGR 20  
DB 29 SHIECRKPKYKELRLEVGR 48

RESULT 2  
ID O98VY6 PRELIMINARY; PRT; 53 AA.  
AC O98VY6;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URU17;  
RA Costa-Mattoli M., Ferre V., Monpoeho S., Garcia L., Collina R.,  
Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis A virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ309228; CAC37074.1; -  
FT NON\_TER 1 53  
FT SEQUENCE 53 AA; 6271 MW; 0846AEP4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGR 20  
DB 29 SHIECRKPKYKELRLEVGR 48

RESULT 3  
ID O98VY5 PRELIMINARY; PRT; 53 AA.  
AC O98VY5;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URU13;  
RA Costa-Mattoli M., Ferre V., Monpoeho S., Garcia L., Collina R.,  
Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis A virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ309229; CAC37075.1; -  
FT NON\_TER 1 53  
FT SEQUENCE 53 AA; 6271 MW; 0846AEP4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;

Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGR 20  
DB 29 SHIECRKPKYKELRLEVGR 48

RESULT 4  
ID O98VY4 PRELIMINARY; PRT; 53 AA.  
AC O98VY4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHILE-J;  
RA Costa-Mattoli M., Ferre V., Monpoeho S., Garcia L., Collina R.,  
Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ309230; CAC37076.1; -  
FT NON\_TER 1 53  
FT SEQUENCE 53 AA; 6271 MW; 0846AEP4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGR 20  
DB 29 SHIECRKPKYKELRLEVGR 48

RESULT 5  
ID O98VY3 PRELIMINARY; PRT; 53 AA.  
AC O98VY3;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=URU16;  
RA Costa-Mattoli M., Ferre V., Monpoeho S., Garcia L., Collina R.,  
Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ309231; CAC37077.1; -  
FT NON\_TER 1 53  
FT SEQUENCE 53 AA; 6271 MW; 0846AEP4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2002, 23:55:42 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-39

Perfect score: 107

Sequence: 1 SHIECRKPKYKELRLEVQKOR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

1: SP:SPREMBL.19:\*  
2: SP:Bacteria:\*  
3: SP:Fungi:\*  
4: SP:Human:\*  
5: SP:Invertebrate:\*  
6: SP:Mammal:\*  
7: SP:MHC:\*  
8: SP:Organelle:\*  
9: SP:Phage:\*  
10: SP:Plant:\*  
11: SP:Rodent:\*  
12: SP:Virus:\*  
13: SP:Vertebrate:\*  
14: SP:Unclassified:\*  
15: SP:\_rVirus:\*  
16: SP:\_bacteriap:\*  
17: SP:\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	53	12	098VY7 hepatitis a
2	107	100.0	53	12	098VY6 hepatitis a
3	107	100.0	53	12	098VY5 hepatitis a
4	107	100.0	53	12	098VY4 hepatitis a
5	107	100.0	53	12	098VY3 hepatitis a
6	107	100.0	53	12	098VY2 hepatitis a
7	107	100.0	53	12	098VY1 hepatitis a
8	107	100.0	53	12	098VY0 hepatitis a
9	107	100.0	55	12	099VU8 hepatitis a
10	107	100.0	55	12	099VU4 hepatitis a
11	107	100.0	55	12	099VU3 hepatitis a
12	107	100.0	55	12	099VU2 hepatitis a
13	107	100.0	56	12	099VU1 hepatitis a
14	107	100.0	56	12	099VU0 hepatitis a
15	107	100.0	56	12	099VU9 hepatitis a
16	107	100.0	56	12	099VU7 hepatitis a

## ALIGNMENTS

RESULT: 1	PRELIMINARY:	PRT:	53 AA.
098VY7			
AC 098VY7			
DT 01-JUN-2001 (TREMUR. 17, Created)			
DT 01-JUN-2001 (TREMUR. 17, Last sequence update)			
DT 01-JUN-2001 (TREMUR. 17, Last annotation update)			
DE POLYPROTEIN (FRAGMENT).			
OS Hepatitis A virus.			
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;			
OC Hepatitis A virus.			
OC NCBI_TaxID=12092;			
OX [1]			
RN SEQUENCE FROM N.A.			
RC STRAIN=URU1;			
RA Billadel S., Vega T., Perez-Bercoff R., Cristina J.,			
RT "Genetic variability of hepatitis A virus in South America reveals			
heterogeneity and co-circulation during epidemic outbreaks."			
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL: AJ309227; CAC37073.1; -			
FT NON_TER			
FT NON_TER			
SQ SEQUENCE	53 AA;	6271 MW;	0846AEF4BC397432 CRC64;

**THIS PAGE BLANK (USPTO)**

---



CC -1- FUNCTION: MOST ABUNDANT PROTEIN IN THE VIRION AND AN IMPORTANT  
CC ELEMENT CONFERRING HELICAL SYMMETRY ON THE NUCLEOPROTEIN CORE AS  
CC WELL AS INTERACTING WITH M PROTEIN DURING VIRION FORMATION.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES NUCLEOCAPSID FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J40504; AAB25601.1; -  
DR PIR: J01533; J01533.  
DR Nucleocapsid.  
KW Nucleocapsid.  
SQ SEQUENCE 391 AA; 43557 MW; ABBC2AD5534D3C CRC64;

Query Match 38.3%; Score 41; DB 1; Length 391;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 IECKRPYKELRLLEVKG 18  
||| ||| :| :| :  
Db 129 IESKRSYKKMKLENGE 144

RESULT 24  
NCAP\_BRSLVA STANDARD; PRT; 391 AA.  
AC P22677;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last annotation update)  
DE Nucleocapsid protein.  
GN N.  
OS Bovine respiratory syncytial virus (strain A51908) (BRV).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
OX NCBI\_TaxID=11247;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91082446; PubMed=1984667;  
RA Samal S.K., Zamora M., McPhillips T.H., Mohanty S.B.;  
RT "Molecular cloning and sequence analysis of bovine respiratory  
RT syncytial virus mRNA encoding the major nucleocapsid protein.",  
RL Virology 180:453-456(1991).  
CC -1- FUNCTION: MOST ABUNDANT PROTEIN IN THE VIRION AND AN IMPORTANT  
CC ELEMENT CONFERRING HELICAL SYMMETRY ON THE NUCLEOPROTEIN CORE AS  
CC WELL AS INTERACTING WITH M PROTEIN DURING VIRION FORMATION.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES NUCLEOCAPSID FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M35076; AAA42812.1; -  
DR PIR: A38525; VHMZB4.  
DR Nucleocapsid.  
KW Nucleocapsid.  
SQ SEQUENCE 391 AA; 43445 MW; F6729E09F02E7F8A CRC64;

Query Match 38.3%; Score 41; DB 1; Length 391;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
OY 3 IECKRPYKELRLLEVKG 18  
||| ||| :| :| :  
Db 129 IESKRSYKKMKLENGE 144

Db 129 IESKRSYKKMKLENGE 144

RESULT 25  
NCAP\_BRSLVR STANDARD; PRT; 391 AA.  
AC O65708;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Nucleocapsid protein.  
GN N.  
OS Bovine respiratory syncytial virus (strain Rd94) (BRV).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
OX NCBI\_TaxID=11249;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dessy F., Walravens K., Knott I., Laloux O., Collard A.,  
RA Letesson J.J., Coppe P.;  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: MOST ABUNDANT PROTEIN IN THE VIRION AND AN IMPORTANT  
CC ELEMENT CONFERRING HELICAL SYMMETRY ON THE NUCLEOPROTEIN CORE AS  
CC WELL AS INTERACTING WITH M PROTEIN DURING VIRION FORMATION.  
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES NUCLEOCAPSID FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L27840; AAA42814.1; -  
DR Nucleocapsid.  
KW Nucleocapsid.  
SQ SEQUENCE 391 AA; 43497 MW; B11667AD5534D62D CRC64;

Query Match 38.3%; Score 41; DB 1; Length 391;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
OY 3 IECKRPYKELRLLEVKG 18  
||| ||| :| :| :  
Db 129 IESKRSYKKMKLENGE 144

Search completed: June 16, 2002, 00:10:05  
Job time: 653 sec

```
RESULT 21
SUCC_METUA STANDARD; PRT; 364 AA.
ID SUCC_METUA
AC 057663:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Succinyl-CoA synthetase beta chain (EC 6.2.1.5) (SCS-beta).
GN SUCC OR M0210.
OS Methanococcus jannaschlii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=6888087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschlii."
RT Science 273:1058-1073(1996).
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + succinate + CoA -> ADP + succinyl-CoA +
CC phosphate.
CC -1- SIMILARITY: TO OTHER BETA SUBUNITS OF SUCCINYL-COA SYNTHETASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67477; AAB98195.1; -.
DR HSSP: P07460; 2SCU.
DR TIGR: MJ0210; -.
DR InterPro: IPR003135; ATP-grasp.
DR InterPro: IPR000303; COA_ligase.
DR Pfam: PF02222; ATP-grasp; 1.
DR Pfam: PF00548; ligase-CoA; 1.
DR PROSITE: PS01217; SUCCINYL_COA_LIG_3; 1.
DR Ligase; Tricarboxylic acid cycle; Complete proteome.
KW SEQUENCE 364 AA; 40907 MW; 16937FC54694770C CRC64;
SQ
Query Match 38.3%; Score 41; DB 1; Length 364;
Best Local Similarity 37.9%; Pred. No. 25;
Matches 11; Conservative 4; Mismatches 2; Indels 12; Gaps 2;
OY 2 HIECKRPY-----KELRL---EVCK 18
DB 138 HIDVKKPFLPYIARWIVKEAKLPSEICK 166
RESULT 22
T2M2_METUA STANDARD; PRT; 370 AA.
ID T2M2_METUA
AC Q36844;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type II restriction enzyme MjaII (EC 3.1.21.4) (Endonuclease MjaII)
DE (R.MjaII).
GN MJAII OR MJ1449.
```

```
OS Methanococcus jannaschlii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=6888087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschlii."
RT Science 273:1058-1073(1996).
RL Science 273:1058-1073(1996).
CC [2]
CC CHARACTERIZATION.
CC Zerkler B., Myers P.A., Escalante H., Roberts R.J.;
CC Unpublished observations (XXX-1997).
CC
CC -1- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE GGNC.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC specific double-stranded fragments with terminal 5'-phosphates.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67585; AAB99461.1; -.
DR REBASE: 1222; MjaII.
DR TIGR: MJ1449; -.
DR HydroLase; Endonuclease; Nuclease; Restriction system;
KW Complete proteome.
KW SEQUENCE 370 AA; 43911 MW; 8B4CF6208F914B33 CRC64;
SQ
Query Match 38.3%; Score 41; DB 1; Length 370;
Best Local Similarity 64.3%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SHIECKRPYKELRL 14
DB 313 SEIECAKAKELSL 326
RESULT 23
NCAP_BRVS3
ID NCAP_BRVS3 STANDARD; PRT; 391 AA.
AC P35943;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nucleocapsid protein.
GN N.
OS Bovine respiratory syncytial virus (strain 391-2) (BRS).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=31611;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92341085; PubMed=1634882;
RA Amann V.L., Lerch R.A., Anderson K., Wertz G.W.;
RA "Bovine respiratory syncytial virus nucleocapsid protein: mRNA
RA sequence analysis and expression from recombinant vaccinia virus
RA vectors."
RT J. Gen. Virol. 73:999-1003(1992).
RL
```

DR InterPro: IPR003601; DNATopI\_ATP\_bind.  
DR InterPro: IPR003602; DNATopI\_DNA\_bind.  
DR InterPro: IPR000380; Pro.Topoisomerase.  
DR InterPro: IPR002936; Toprim.  
DR Pfam: PF01131; Topoisom\_bac; 1.  
DR Pfam: PF01751; Toprim; 1.  
DR Pfam: PF01396; 2i-C4\_Topoisom; 2.  
DR PRINTS: PR00417; PRPISMRASE1.  
DR SMART: SM00437; TOP1AC; 1.  
DR SMART: SM00436; TOP1BC; 1.  
DR SMART: SM00493; TOPRIM; 1.  
DR PROSITE: PS00396; TOPOISOMERASE\_I\_PROK; 1.  
DR Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
KW Repeat; Complete proteome.  
FT 2N-FING 596 628 C4-TYPE 1.  
FT 2N-FING 658 685 C4-TYPE 2.  
FT 2N-FING 707 732 C4-TYPE 3.  
FT ACT\_SITE 318 318 DNA\_CLEAVAGE (BY SIMILARITY).  
SQ SEQUENCE 861 AA; 99739 MW; 903BFEA5265599 CRC64;

Query Match 39.3%; Score 42; DB 1; Length 861;  
Best Local Similarity 40.0%; Pred. No. 40;  
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 IECKRKYKRLRLEVG 17  
|:|:|:|:|:|:|:  
Db 705 IQCEKCYNMKTKIG 719

RESULT 19  
END4\_BACHD STANDARD; PRT; 298 AA.  
AC O9KD33;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable endonuclease IV (EC 3.1.21.2) (Endodeoxyribonuclease IV).  
GN NFO OR BH1386.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
CC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-125 / JCM 9153;  
RX MEDLINE=20512582; Pubmed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Ogasawara N., Kunara S.,  
RA Fujii F., Hitama C., Nakamura Y.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -1- FUNCTION: Endonuclease IV plays a role in DNA repair. It cleaves  
phosphodiester bonds at apurinic or apyrimidinic sites (AP sites)  
to produce new 5' ends that are base-free deoxyribose 5-phosphate  
residues. It preferentially attacks modified AP sites created by  
bleomycin and neocarzinostatin (By similarity).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
phosphooligonucleotide end-products.  
CC -1- COFACTOR: Binds 3 zinc ions (By similarity).  
CC -1- SIMILARITY: BELONGS TO AP ENDONUCLEASES FAMILY 2.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AP001511; BAB05105.1; -  
DR HSSP: P12638; 1QTW.

DR InterPro: IPR001719; AP\_endonuclease-2.  
DR Pfam: PF01261; AP\_endonuclease2; 1.  
DR SMART: SM00518; AP2EC; 1.  
DR PROSITE: PS00729; AP\_NUCLEASE\_F2\_1; 1.  
DR PROSITE: PS00730; AP\_NUCLEASE\_F2\_2; 1.  
DR PROSITE: PS00731; AP\_NUCLEASE\_F2\_3; 1.  
KW Hydroxylase; Nuclease; Endonuclease; DNA repair; Metal-binding; Zinc;  
Complete proteome.  
FT METAL 70 70 ZINC 1 (BY SIMILARITY).  
FT METAL 111 111 ZINC 1 AND 2 (BY SIMILARITY).  
FT METAL 146 146 ZINC 1 AND 2 (BY SIMILARITY).  
FT METAL 180 180 ZINC 2 (BY SIMILARITY).  
FT METAL 183 183 ZINC 3 (BY SIMILARITY).  
FT METAL 215 215 ZINC 2 (BY SIMILARITY).  
FT METAL 228 228 ZINC 3 (BY SIMILARITY).  
FT METAL 230 230 ZINC 3 (BY SIMILARITY).  
FT METAL 260 260 ZINC 2 (BY SIMILARITY).  
FT METAL 260 260 ZINC 2 (BY SIMILARITY).  
SQ SEQUENCE 298 AA; 32997 MW; 9A25ED81A755B4B CRC64;

Query Match 38.3%; Score 41; DB 1; Length 298;  
Best Local Similarity 53.8%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 6 RKPYKRLRLEVGK 18  
|:|:|:|:|:|:|:  
Db 44 RKPIELNIEAGR 56

RESULT 20  
URED\_SYNPV STANDARD; PRT; 319 AA.  
AC O87399;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Urease accessory protein ured.  
GN URED.  
OS Synechococcus sp. (strain WH7805).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=59931;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Collier J.L., Brahamsa B., Palenik B.;  
RT "Molecular genetic and biochemical characterization of urease (urea  
amidohydrolase, EC 3.5.1.5) from the marine cyanobacterium,  
Synechococcus WH7805.";  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PROBABLY FACILITATES NICKEL INCORPORATION.  
CC -1- SIMILARITY: BELONGS TO THE URED FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF056189; AAC61499.1; -  
DR InterPro: IPR002669; Ured.  
DR Pfam: PF01774; Ured; 1.  
KW Nickel.  
SQ SEQUENCE 319 AA; 35401 MW; 0E491227E8E4771A CRC64;

Query Match 38.3%; Score 41; DB 1; Length 319;  
Best Local Similarity 57.1%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
OY 5 CRKPYKRLRLEVGK 18  
|:|:|:|:|:|:|:  
Db 33 CSAPEKRLRAEGK 46

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X51680; CAA35983.1; -

DR PIR: S07532; S07532.  
KW Signal; Colled coil; Glycoprotein.  
FT SIGNAL 19 OR 21 (POTENTIAL).  
FT CHAIN 20 286 PUFF II/9-1 PROTEIN.  
FT DOMAIN 61 235 HELICAL (POTENTIAL).  
FT CARBOHD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 286 AA; 32034 MW; AAG7B55F191BBID CRC64;

Query Match 39.3%; Score 42; DB 1; Length 286;  
Best Local Similarity 57.1%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 IECRRPYKELRELV 16  
DB 195 IACRKGQELRCEI 208

RESULT 17  
SYM\_PYRHO STANDARD; PRT; 723 AA.  
AC 058721;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)  
DE (MEERS).  
GN MERS OR PH0993.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=96344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.,  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING  
REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO  
FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR  
TRNA(ENER) AMINOACYLATION.  
CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) -> AMP +  
diphosphate + L-methionyl-tRNA(Met).  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC STRONG, TO CYSTEINYL-TRNA SYNTHETASE.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AP000004; BAA30090.1; -  
HSP: P00959; IMED.

InterPro: IPR002300; tRNA-synt\_1a.  
DR InterPro: IPR001412; tRNA-synt\_1.  
DR InterPro: IPR002304; tRNA-synt\_mec.  
DR InterPro: IPR002547; tRNA\_bind.  
DR Pfam: PF00133; tRNA-synt\_1; 1.  
DR Pfam: PF01588; tRNA\_bind; 1.  
DR Pfam: PF01041; TRANSTHET.  
DR PROSITE: PS00178; AA-TRNA-LIGASE\_I; 1.  
KW Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding;  
KW tRNA-binding; Metal-binding; zinc; Complete proteome.  
FT SITE 11 21 "HIGH" REGION.  
FT SITE 344 348 "KMSK" REGION.  
FT DOMAIN 569 628 LINKER.  
FT DOMAIN 629 721 TRNA BINDING.  
FT METAL 143 143 ZINC (BY SIMILARITY).  
FT METAL 146 146 ZINC (BY SIMILARITY).  
FT METAL 156 156 ZINC (BY SIMILARITY).  
FT METAL 159 159 ZINC (BY SIMILARITY).  
SQ SEQUENCE 723 AA; 84743 MW; F839FDBCE7DC09DC CRC64;

Query Match 39.3%; Score 42; DB 1; Length 723;  
Best Local Similarity 42.1%; Pred. No. 34;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 SHIECRPYKELREYKQ 19  
DB 594 SKILLKRYKDIKEGKE 612

RESULT 18  
TOP1\_BUCAI STANDARD; PRT; 861 AA.  
AC P57371;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
DE (unwisting enzyme) (Swivelase).  
GN TOPA OR BU284.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
synbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=18099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TOKYO 1998;  
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.,  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
Buchnera sp. APS.";  
RL Nature 407:81-86(2000).  
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
DNA, followed by passage and rejoining.  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AP001118; BAB12994.1; -

OX NCBI\_TaxID=11251;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9279331; PubMed=2525176;  
 RA Johnson P.R., Collins P.L.;  
 RT "The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial  
 virus (RSV) of antigenic subgroups A and B: sequence conservation and  
 divergence within RSV genomic RNA."  
 RL J. Gen. Virol. 70:1539-1547(1989).  
 CC -1- FUNCTION: MOST ABUNDANT PROTEIN IN THE VIRION AND AN IMPORTANT  
 ELEMENT CONFERRING HELICAL SYMMETRY ON THE NUCLEOPROTEIN CORE AS  
 WELL AS INTERACTING WITH M PROTEIN DURING VIRION FORMATION.  
 CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES NUCLEOCAPSID FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D00736; BAA00637.1; -  
 DR PIR: C32063; VHN23.  
 KW Nucleocapsid.  
 SQ SEQUENCE 391 AA; 43416 MW; 9A2BBB50103B2835 CRC64;  
 QY 3 ECRKPKRELRLVGVK 18  
 Db 129 IESRSYKLLKEKME 144  
 |||||:|:|:  
 RESULT 15  
 ID LOLA\_XYLFA STANDARD; PRT; 210 AA.  
 AC 09PDC7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Outer-membrane lipoproteins carrier protein precursor.  
 GN LOLA OR XF1452.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,  
 Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin L., Bove J.M., Briones M.R.S.,  
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 Falcinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 Fraja J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 Granger M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 Rana Nihal A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Ouaguo R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva H.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Trufil D., Tsai S.M., Tsunako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."  
 RL Nature 406:151-159(2000).  
 CC -1- FUNCTION: PARTICIPATES IN THE TRANSLOCATION OF LIPOPROTEINS FROM  
 THE INNER MEMBRANE TO THE OUTER MEMBRANE. ONLY FORMS A COMPLEX  
 WITH A LIPOPROTEIN IF THE RESIDUE AFTER THE N-TERMINAL CYS IS NOT  
 AN ASPARTATE (THE ASP ACTS AS A TARGETING SIGNAL TO INDICATE THAT  
 THE LIPOPROTEIN SHOULD STAY IN THE INNER MEMBRANE) (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE LOLA FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE003975; AAF84261.1; ALT. INIT.  
 DR Chapterone; Transport; Protein transport; Periplasmic; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 210 OUTER-MEMBRANE LIPOPROTEINS CARRIER  
 FT PROTEIN.  
 SQ SEQUENCE 210 AA; 23618 MW; 08B8BBA44005B24F CRC64;  
 QY 4 ECRKPKRELRLVGVK 19  
 Db 73 ECDAPYKOLVYADGKR 88  
 |||||:|:|:  
 RESULT 16  
 ID P091\_SCICO STANDARD; PRT; 286 AA.  
 AC P22311;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE PUFF II/9-1 protein precursor.  
 GN II/9-1.  
 OS Sclara coprophila (Fungus gnat).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Scleroidea;  
 OC Scleridae; Bradysia.  
 OX NCBI\_TaxID=38358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6980;  
 RX MEDLINE=90133907; PubMed=2614832;  
 RA DiPaccolomeis S.M., Gerbl S.A.;  
 RT "Molecular characterization of DNA puff II/9A genes in Sclara  
 coprophila."  
 RL J. Mol. Biol. 210:531-540(1989).  
 CC -1- MISCELLANEOUS: THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED  
 INTERMOLECULAR COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR  
 DISULFIDE BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION  
 D OF THE HEPTAD REPEAT.  
 CC -1- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-2 PROTEIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

DR InterPro: IPR001680; WD40.  
DR SMART: SM00320; WD40; 3.  
KW Hypothetical protein.  
SQ SEQUENCE 420 AA; 47525 MW; 9DC1F710FBEB9761 CRC64;

Query Match 43.0%; Score 46; DB 1; Length 420;  
Best Local Similarity 46.7%; Pred. No. 4.4;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 4 BCRKPKYKELRLEVQK 18  
DB 376 DCSLPFKERVDGK 390

RESULT 13  
CO3\_RAT  
ID CO3\_RAT STANDARD: PRT; 1663 AA.  
AC P01026;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Complement C3 precursor [contains: C3a anaphylatoxin].  
GN C3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
ON NCBI\_TaxID=10116;  
RX STRAIN=WISTAR; TISSUE=Liver;  
RX MEDLINE=90245672; PubMed=2336397;  
RA Miumi Y., Sohma M., Ikehara Y.;  
RT "Nucleotide and deduced amino acid sequence of rat complement C3.";  
RL Nucleic Acids Res. 18:2178-2178(1990).  
RN  
RP SEQUENCE OF 671-748.  
RX MEDLINE=79062262; PubMed=309768;  
RA Jacobs J.W., Rubin J.S., Hugli T.E., Bogardt R.A., Mariz I.R.,  
RA Daniels J.S., Daughaday W.H., Bradshaw R.A.;  
RT "Purification, characterization, and amino acid sequence of rat  
anaphylatoxin (C3a).";  
RL Biochemistry 17:5031-5038(1978).  
RN  
RP SEQUENCE OF 1316-1595 FROM N.A.  
RX MEDLINE=89380332; PubMed=2674144;  
RA Sundstrom S.A., Komm B.S., Ponce-De-Leon H., Yi Z., Teuscher C.,  
RA Lyttle C.R.;  
RT "Estrogen regulation of tissue-specific expression of complement C3.";  
RL J. Biol. Chem. 264:16941-16947(1989).  
CC -1- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE  
COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL  
REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.  
CC AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE  
THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.  
CC C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
BASOPHILIC LEUKOCYTES.  
CC -1- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG  
RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE  
BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,  
RELEASES C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA  
CHAIN).  
CC -1- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.  
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: X52477; CAA36716.1; -.  
DR EMBL: M29866; AAA40837.1; ALT\_SEQ.  
DR PIR: A01260; A01260.  
DR PIR: S15764; S15764.  
DR HSSP: P01024; 1C3D.  
DR InterPro: IPR002890; A2M\_N.  
DR InterPro: IPR001599; Alpha\_2\_macroloblin.  
DR InterPro: IPR000020; Anaphylatoxin.  
DR InterPro: IPR001840; Anaphylatoxin.  
DR InterPro: IPR001134; Netrin\_C.  
DR Pfam: PF00207; A2M\_N; 1.  
DR Pfam: PF01835; A2M\_N; 1.  
DR Pfam: PF01821; ANATO; 1.  
DR Pfam: PF01759; NTR; 1.  
DR PRINTS: PR00004; ANAPHYLATOXN.  
DR ProDom: PD003264; Anaphylatoxin; 1.  
DR SMART: SM00104; ANATO; 1.  
DR PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
DR PROSITE: PS01177; ANAPHYLATOXIN\_1; 1.  
DR PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.  
KM Complement pathway; Complement alternate pathway; Plasma;  
KM Inflammatory response; Glycoprotein; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 1663 COMPLEMENT C3.  
FT CHAIN 25 666 BETA CHAIN.  
FT CHAIN 671 1663 ALPHA CHAIN.  
FT PEPTIDE 671 748 C3A ANAPHYLATOXIN.  
FT CHAIN 749 1663 C3B (ALPHA CHAIN).  
FT SITE 748 749 CLEAVAGE (BY C3 CONVERTASE).  
FT DOMAIN 693 728 ANAPHYLATOXIN-LIKE.  
FT DISULFID 558 816 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 626 661 BY SIMILARITY.  
FT DISULFID 693 720 BY SIMILARITY.  
FT DISULFID 694 727 BY SIMILARITY.  
FT DISULFID 707 728 BY SIMILARITY.  
FT DISULFID 873 1513 BY SIMILARITY.  
FT DISULFID 1101 1158 BY SIMILARITY.  
FT DISULFID 1358 1489 BY SIMILARITY.  
FT DISULFID 1389 1458 BY SIMILARITY.  
FT DISULFID 1506 1511 BY SIMILARITY.  
FT DISULFID 1518 1590 BY SIMILARITY.  
FT DISULFID 1537 1661 BY SIMILARITY.  
FT THIOLEST 1010 1013 BY SIMILARITY.  
FT CARBOHYD 939 939 N-LINKED (GICNAC. . .) (PROBABLE).  
FT CARBOHYD 1617 1617 N-LINKED (GICNAC. . .) (PROBABLE).  
FT CONFLICT 721 722 LK -> KL (IN REF. 2).  
SQ SEQUENCE 1663 AA; 186460 MW; 2F87CB143CDD4BC CRC64;

Query Match 40.7%; Score 43.5; DB 1; Length 1663;  
Best Local Similarity 47.4%; Pred. No. 4.4;  
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

OY 1 SHIECKPKYKELRLEVQK 19  
DB 1586 SHVKCR---NALKLOKQK 1601

RESULT 14  
NCAP\_HRSVL  
ID NCAP\_HRSVL STANDARD: PRT; 391 AA.  
AC P24566;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last annotation update)  
DE Nucleocapsid protein.  
GN N.  
OS Human respiratory syncytial virus (subgroup B / strain 18537).  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.

RX MEDLINE-91311420; PubMed-1649901;  
RA Tsarev S.A., Emerson S.O., Balayan M.S., Ticehurst J.R.,  
RA Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE-9232168; PubMed-2541023;  
RA Balayan M.S., Kusov Y.Y., Andjapardize A.G., Tsarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
simian hepatitis A viruses";  
RL FEBS Lett. 247:425-428(1989).  
CC -1- SUBMIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: D00924; BAA00766.1; -;  
DR EMBL: X15461; CAA33490.1; -;  
DR PIR: A30470; GNNYSA.  
DR PIR: S04885; S04885.  
DR MEROPS: C03.005; -;  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease;  
FT CHAIN 1 27  
FT CHAIN 28 249 COAT PROTEIN VP4 (PIA).  
FT CHAIN 250 495 COAT PROTEIN VP2 (PIB).  
FT CHAIN 496 795 COAT PROTEIN VP3 (PIC).  
FT CHAIN 796 984 COAT PROTEIN VP1 (PID).  
FT CHAIN 985 1091 CORE PROTEIN P2A.  
FT CHAIN 1092 1426 CORE PROTEIN P2B.  
FT CHAIN 1427 1498 CORE PROTEIN P2C.  
FT CHAIN 1499 1521 PROBABLE PROTEIN 3A.  
FT CHAIN 1522 1741 PROBABLE PROTEIN 3B.  
FT CHAIN 1742 2230 PROBABLE PROTEIN 3C.  
FT CHAIN 2230 231296 RNA-DIRECTED POLYMERASE 3D.  
FT CHAIN 2230 231296 MW; 87B3230E324E1F19 CRC64;  
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;  
-----  
Query Match 80.4%; Score 86; DB 1; Length 2230;  
Best Local Similarity 85.0%; Pred. No. 6.9e-06;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
-----  
QY 1 SHIECRKPKYKELREVGKOR 20  
DB 796 SHIEGRKPKYKELREVGKOR 815  
-----  
RESULT 11  
POLG\_HPAVT STANDARD; PRT; 839 AA.  
AC P31788;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
P2A] (Fragment).  
OS Simian hepatitis A virus (strain CY-145).  
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.  
OX NCBI\_TaxID-31707;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91311421; PubMed-1649902;  
RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
RT "Sequence analysis of a new hepatitis A virus naturally infecting  
cynomolgus macaques (Macaca fascicularis)".;  
RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBMIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M59286; AAA45473.1; -;  
DR PIR: J01180; GNNYS2.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245 COAT PROTEIN VP4 (PIA).  
FT CHAIN 246 491 COAT PROTEIN VP2 (PIB).  
FT CHAIN 492 7 COAT PROTEIN VP3 (PIC).  
FT CHAIN 7 >839 COAT PROTEIN VP1 (PID).  
FT CHAIN 839 839 CORE PROTEIN P2A.  
FT NON\_TER 839 839  
SQ SEQUENCE 839 AA; 93825 MW; 2CACCA4BD1E192DBC CRC64;  
-----  
Query Match 79.0%; Score 84.5; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 4.6e-06;  
Matches 18; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
-----  
QY 1 SHIECRKPKYKELREVGKOR 20  
DB 792 SHIE-KKPKYKELREVGKOR 810  
-----  
RESULT 12  
YAGA\_SCHPO STANDARD; PRT; 420 AA.  
AC 009873;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Hypothetical 47.5 kDa protein C12G12.10 in chromosome I.  
GN SPAC12G12.10.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID-4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA Devlin K., Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,  
RA Walsh S.V.;  
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: Z66568; CAA91505.1; -;

Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRELVGKOR 20  
|||||  
Db 792 SHIECRKPKYKELRELVGKOR 811

RESULT 8  
POLG\_HPAVM STANDARD; PRT: 2227 AA.  
ID POLG\_HPAVM 081083; 081084; 081085; 081086; 081087; 081088; 081089;  
AC 081090; 081091; 081092; 081093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MBB).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Missel T., Klehn R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A virus (isolate MBB)."  
RL Virus Res. 8:153-171(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M20773; AAA45474.1; -.  
DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRELVGKOR 20

|||||  
Db 792 SHIECRKPKYKELRELVGKOR 811

RESULT 9  
POLG\_HPAVG STANDARD; PRT: 808 AA.  
ID POLG\_HPAVG 002381;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
OS Hepatitis A virus (strain GA76).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=31706;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92260183; PubMed=1316423;  
RA Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;  
RT "Characterization of a genetic variant of human hepatitis A virus.";  
RL J. Med. Virol. 36:118-124(1992).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M66695; AAA45477.1; -.  
DR Polyprotein; Coat protein; Core protein.  
KW NON\_TER 1  
FT CHAIN 1 2  
FT CHAIN 3 223  
FT CHAIN 224 470  
FT CHAIN 471 770  
FT CHAIN 771 >808  
FT CHAIN 808 808  
SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 88.8%; Score 95; DB 1; Length 808;  
Best Local Similarity 95.0%; Pred. No. 8.7e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRELVGKOR 20  
|||||  
Db 771 SHIECRKPKYKELRELVGKOR 790

RESULT 10  
POLG\_HPAVS STANDARD; PRT: 2230 AA.  
ID POLG\_HPAVS 014553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12102;  
RN [1]  
RP SEQUENCE FROM N.A.



RT comparison with different strains of hepatitis A virus and other  
RT Picornaviruses.";  
RL J. Virol. 61:50-59(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATTENUATED;  
RA MEDLINE-81175701; PubMed-3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,  
RA Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RT comparison with wild-type virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RN [3]  
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE-85166289; PubMed-2984684;  
RA Barclay B.M., Ticehurst J.R., Miele T.A., Matzel J.V. Jr.,  
RA Purcell R.H., Feinstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RT proteins and RNA polymerase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: STRAIN HM-175/77 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between The Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M14114; AAA45475.1; -;  
DR EMBL; M14707; AAA45465.1; -;  
DR EMBL; M14707; AAA45466.1; -; ALT\_INIT.  
DR EMBL; M16632; AAA45471.1; -;  
DR PIR; A25981; GNNYHM.  
DR PIR; A25914; GNNYMK.  
DR PIR; A03905; A03905.  
DR MEROPS; C03.005; -;  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; RNA polymerase; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT CHAIN 2227 77  
FT CHAIN 77 77  
FT CHAIN 764 764  
FT CHAIN 821 821  
FT CHAIN 1052 1052  
FT CHAIN 1062 1062  
FT CHAIN 1118 1118  
FT CHAIN 1151 1151  
FT CHAIN 1163 1163  
FT CHAIN 1277 1277  
FT CHAIN 1500 1500  
FT VARIANT H -> Y (IN ATTENUATED STRAIN).  
FT VARIANT

FT VARIANT 1805 1805 D -> N (IN ATTENUATED STRAIN).  
FT VARIANT 1930 1930 S -> T (IN ATTENUATED STRAIN).  
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AE8740A6 CRC64;  
Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2,6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 SHIECRKPKKLEVGKOR 20  
Db 792 SHIECRKPKKLEVGKOR 811  
RESULT 7  
POLG\_HP AVL STANDARD; PRT; 2227 AA.  
ID POLG\_HP AVL  
AC P06441:  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 to VP4; Core proteins  
DE P2A to P2C; Probable proteins P3A to P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-85190549; PubMed-2986127;  
RA Nejlarian R., Caput D., Gee W.M., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between The Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; K02990; AAA45472.1; -;  
DR PIR; A03903; GNNYHR.  
DR MEROPS; C03.005; -;  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1076  
FT CHAIN 1077 1422  
FT CHAIN 1423 1484  
FT CHAIN 1485 1507  
FT CHAIN 1508 1678  
FT CHAIN 1679 2227  
FT CHAIN 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;  
SQ SEQUENCE

Query Match

100.0%; Score 107; DB 1; Length 2227;

OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
variants arising during persistent infection: evidence for genetic  
recombination."  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M59809; AAA45469.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease;  
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;  
  
Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 SHICRKPKELELVGKOR 20  
Db 792 SHICRKPKELELVGKOR 811  
|||||  
RESULT 5  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
ID POLG\_HPAV8  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 186).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12096;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE-91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
variants arising during persistent infection: evidence for genetic  
recombination."  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M59808; AAA45467.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease;  
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;  
  
Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 SHICRKPKELELVGKOR 20  
Db 792 SHICRKPKELELVGKOR 811  
|||||  
RESULT 6  
POLG\_HPAV8 STANDARD; PRT; 2227 AA.  
ID POLG\_HPAV8  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H1D TYPE;  
RX MEDLINE-87061253; PubMed=3023706;  
RA Cohen J.T., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus:

SO SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;

Query Match 100.0%; Score 107; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 4,1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELREVGKOR 20  
|  
Db 296 SHIECRKPYKELREVGKOR 315

RESULT 2  
POLG\_HPAVC STANDARD; PRT; 852 AA.

AC P06442; Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
OS Hepatitis A virus (strain CR326).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85185648; Pubmed=2985793;  
RA Lhennery D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
RA Young A., Mitra S.W.;  
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA,"  
J. Virol. 54:247-255(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M10033; AAA45470.1; -  
DR PIR; A03904; GNNYHA.  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 >852  
FT NON\_TER 852  
SQ SEQUENCE 852 AA; 95563 MW; 73D3BD0AD532820E CRC64;

Query Match 100.0%; Score 107; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELREVGKOR 20  
|  
Db 792 SHIECRKPYKELREVGKOR 811

RESULT 3  
POLG\_HPAVC STANDARD; PRT; 2226 AA.

AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; Pubmed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M59810; AAA45468.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6BAE2BF CRC64;

Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELREVGKOR 20  
|  
Db 792 SHIECRKPYKELREVGKOR 811

RESULT 4  
POLG\_HPAVC STANDARD; PRT; 2226 AA.

AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 43c).

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2002, 23:59:12 ; Search time 59.43 Seconds

(without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432a-39

Perfect score: 107

Sequence: 1 SHIECKRKYKELRLEVGKOR 20

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Swissprot\_40:\*

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	341	1	POLG_HPAV1
2	107	100.0	852	1	POLG_HPAVC
3	107	100.0	2226	1	POLG_HPAV2
4	107	100.0	2226	1	POLG_HPAV4
5	107	100.0	2226	1	POLG_HPAV8
6	107	100.0	2227	1	POLG_HPAVH
7	107	100.0	2227	1	POLG_HPAVL
8	107	100.0	2227	1	POLG_HPAVM
9	95	88.8	808	1	POLG_HPAVG
10	86	80.4	2230	1	POLG_HPAVS
11	84.5	79.0	839	1	POLG_HPAVT
12	46	43.0	420	1	YAGA_SCHHO
13	43.5	40.7	1663	1	CO3_RAT
14	43	40.2	391	1	NCAP_HRSY1
15	42	39.3	210	1	LOLA_XYLPA
16	42	39.3	286	1	SYM_PYRHO
17	42	39.3	723	1	SYM_SCICO
18	42	39.3	861	1	TOPI_BUCAI
19	41	38.3	298	1	EMD4_BACHD
20	41	38.3	319	1	URED_SYNPV
21	41	38.3	364	1	SUCC_METVA
22	41	38.3	370	1	T2M2_METVA
23	41	38.3	391	1	NCAP_BRSV3
24	41	38.3	391	1	NCAP_BRSV4
25	41	38.3	391	1	NCAP_BRSVR
26	41	38.3	391	1	NCAP_HRSYA
27	41	38.3	492	1	ANKH_HUMAN
28	41	38.3	492	1	ANKH_MOUSE
29	41	38.3	492	1	ANKH_RAT
30	41	38.3	492	1	ANKH_XENLA
31	41	38.3	4544	1	LRPI_CHICK
32	41	38.3	4544	1	LRPI_HUMAN
33	40.5	37.9	141	1	V16K_TRTVC

34	40.5	37.9	1120	1	RPOM_SCHPO
35	40.5	37.9	1663	1	CO3_MOUSE
36	40.5	37.9	2198	1	YLJ2_CAEEL
37	40	37.4	69	1	MOP_HAEIN
38	40	37.4	309	1	US16_HCMVA
39	40	37.4	345	1	EBB1_MOUSE
40	40	37.4	391	1	NCAP_ORSWA
41	40	37.4	458	1	NIFK_CLOPA
42	40	37.4	558	1	V369_MYCPN
43	40	37.4	1017	1	DPOL_ADEB2
44	40	37.4	3432	1	POLG_JAEV1
45	40	37.4	3432	1	POLG_JAEV5
46	40	37.4	3432	1	POLG_JAEVJ
47	39.5	36.9	422	1	POLG_RV2
48	39	36.4	141	1	GAT3_YEAST
49	39	36.4	177	1	CYSC_SYNP3
50	39	36.4	236	1	VPAD_BPP22
51	39	36.4	376	1	HOSC_THETH
52	39	36.4	406	1	YG43_YEAST
53	39	36.4	472	1	PEDA_STRPY
54	39	36.4	769	1	ITB2_BOVIN
55	39	36.4	859	1	TLR5_MOUSE
56	39	36.4	1226	1	METH_ECOLI
57	39	36.4	1226	1	METH_SALTY
58	39	36.4	1235	1	DPOL_HSV11
59	39	36.4	1235	1	DPOL_HSV1A
60	39	36.4	1235	1	DPOL_HSV1K

## ALIGNMENTS

RESULT	ID	POLG_HPAV1	STANDARD:	PRT:	341 AA.
AC	P13672	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Genome polyprotein [Contains: Coat proteins VP1 to VP3; Core protein P2a] (Fragment).				
OS	Hepatitis A virus (strain LCDC-1).				
OC	Virus; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;				
CC	Hepatitis A virus (HAV)."				
OX	NCBI_TaxID=12093;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89263805; Pubmed=2542903;				
RT	Andonov A.P., Lau P., Chaudhary R.;				
RT	"Nucleotide sequence of the VP1 gene from a Chinese strain of				
RL	Hepatitis A virus (HAV)."				
CC	Nucleic Acids Res. 17:3594-3594(1989).				
CC	-1' SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,				
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,				
CC	VP3, AND VP4.				
CC	-1' PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; X14666; CAA32794.1; -				
DR	PIR; S04137; S04137.				
KW	Polyprotein; Coat protein; Core protein.				
FT	NON_TER	1			
FT	CHAIN	<1	1		COAT PROTEIN VP3 (1C).
FT	CHAIN	2	340		COAT PROTEIN VP1 (1D).
FT	CHAIN	341	>341		CORE PROTEIN P2A.
FT	NON_TER	341	341		



```

; APPLICANT: Straub, Volker
; TITLE OF INVENTION: K-SARCOGLYCAN NUCLEIC ACID SEQUENCES, AMINO
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/119,827
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF-9601
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 256 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-119-827-2

```

```

Query Match      35.5%; Score 38; DB 4; Length 256;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 8 PYKELRL 15
|:|||||
Db 187 PYKELRL 194

```

```

RESULT 25
; Sequence 2, Application US/08651940
; Patent No. 594885
; GENERAL INFORMATION:
; APPLICANT: Stein, Bernd
; APPLICANT: Young, Maria X.H.
; APPLICANT: Young, David B.
; TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE P38-2
; TITLE OF INVENTION: AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,940
; FILING DATE: 20-MAY-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.412
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-651-940-2

```

```

Query Match      35.5%; Score 38; DB 2; Length 364;
Best Local Similarity 57.1%; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 SHIECRPYKELRL 14
|:|||||
Db 61 SLIHARRYRELRL 74

```

Search completed: June 16, 2002, 00:03:13  
Job time: 8527 sec

RESULT 21  
US-09-357-251-22  
; Sequence 22, Application US/09357251  
; Patent No. 6271441  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Farnodu, Lajo O.  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Schwaber, James S.  
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase  
; FILE REFERENCE: BB-1193  
; CURRENT APPLICATION NUMBER: US/09/357,251  
; CURRENT FILING DATE: 1999-07-20  
; EARLIER APPLICATION NUMBER: 60/093,530  
; EARLIER FILING DATE: July 21, 1998  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 22  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-357-251-22

Query Match 35.5%; Score 38; DB 4; Length 123;  
Best Local Similarity 46.7%; Pred. No. 27;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 IECKRKYKLELVG 17  
Db 34 LKCRMKKEIFLEMG 48

RESULT 22  
US-08-651-940-4  
; Sequence 4, Application US/08651940  
; Patent No. 5948885  
; GENERAL INFORMATION:  
; APPLICANT: Stein, Bernd  
; APPLICANT: Yang, Maria X.H.  
; APPLICANT: Young, David B.  
; TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE p38-2  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,940  
; FILING DATE: 20-MAY-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 860098.412  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 155 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-651-940-4

Query Match 35.5%; Score 38; DB 2; Length 155;  
Best Local Similarity 57.1%; Pred. No. 35;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 SHIECKRKYKLERL 14  
Db 61 SLIHARTYRELRL 74

RESULT 23  
US-08-719-758-2  
; Sequence 2, Application US/08719758  
; Patent No. 5837537  
; GENERAL INFORMATION:  
; APPLICANT: Campbell, Kevin P.  
; APPLICANT: Jung, Daniel  
; APPLICANT: Duclos, Franck  
; APPLICANT: Straub, Volker  
; TITLE OF INVENTION: K-SARCOGLYCAN NUCLEIC ACID SEQUENCES, AMINO  
; TITLE OF INVENTION: ACID SEQUENCES AND APPLICATIONS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: USA  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,758  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: UIRF-9601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 256 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-719-758-2

Query Match 35.5%; Score 38; DB 2; Length 256;  
Best Local Similarity 87.5%; Pred. No. 61;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 8 PYKELRL 15  
Db 187 PFKELRL 194

RESULT 24  
US-09-119-827-2  
; Sequence 2, Application US/09119827  
; Patent No. 6211340  
; GENERAL INFORMATION:  
; APPLICANT: Campbell, Kevin P.  
; APPLICANT: Jung, Daniel  
; APPLICANT: Duclos, Franck

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,770B  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, M. Lisa  
REGISTRATION NUMBER: 34,045  
REFERENCE/DOCKET NUMBER: 2002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516)273-2828  
TELEFAX: (516)273-1717  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-015-770B-26

Query Match 35.5%; Score 38; DB 1; Length 31;  
Best Local Similarity 40.0%; Pred. No. 6;  
Matches 8; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Oy 3 IECKRPYKELR--LEVGKOR 20  
| | : | | : | | : | |  
Db 5 INCTRPYNVRRSLSIGPR 24

RESULT 19  
US-08-488-252-15  
Sequence 15, Application US/08488252  
Patent No. 5763160  
GENERAL INFORMATION:  
APPLICANT: Chang Y1 Wang  
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND PROCESS  
TITLE OF INVENTION: OF USING SAME FOR THE DETECTION OF ANTIBODIES TO  
TITLE OF INVENTION: HUMAN IMMUNODEFICIENCY VIRUS (HIV) GP120 ENVELOPE  
TITLE OF INVENTION: PROTEIN, DIAGNOSIS OF AIDS AND PRE-AIDS CONDITIONS  
TITLE OF INVENTION: AND AS VACCINES  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVE.  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,252  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08\326,676  
FILING DATE: 07-Jun-1995  
APPLICATION NUMBER: 07\726,605  
FILING DATE: 09-July-1991  
APPLICATION NUMBER: 07\663,262  
FILING DATE: 01-Mar-1991  
APPLICATION NUMBER: 07\155,321  
FILING DATE: 12-Feb-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C. H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4004 USA  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: Amino acid  
STRANDEDNESS:  
TOPOLOGY: Unknown  
US-08-488-252-15

Query Match 35.5%; Score 38; DB 1; Length 31;  
Best Local Similarity 40.0%; Pred. No. 6;  
Matches 8; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Oy 3 IECKRPYKELR--LEVGKOR 20  
| | : | | : | | : | |  
Db 5 INCTRPYNVRRSLSIGPR 24

RESULT 20  
US-08-651-940-3  
Sequence 3, Application US/08651940  
Patent No. 594885  
GENERAL INFORMATION:

APPLICANT: Steiu, Bernd  
APPLICANT: Yang, Maria X.H.  
APPLICANT: Young, David B.  
TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE p38-2  
TITLE OF INVENTION: AND METHODS OF USE THEREFOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,940  
FILING DATE: 20-MAY-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 860098.412  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 102 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-651-940-3

Query Match 35.5%; Score 38; DB 2; Length 102;  
Best Local Similarity 57.18; Pred. No. 22;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 SHIECKRPYKELRL 14  
| | : | | : | | : | |  
Db 61 SLIHARRTYRELRL 74



```

Query Match          36.4%; Score 39; DB 1; Length 24;
Best Local Similarity 33.3%; Pred. No. 3;
Matches 6; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 3 IECKRPYKELREVGKOR 20
    | : | | : | : | : |
    5 INCTRPYNVNRSLRSTR 22

RESULT 15
US-09-393-554-2
; Sequence 2, Application US/09393554
; Patent No. 6210897
; GENERAL INFORMATION:
; APPLICANT: Andersson, Lelf
; APPLICANT: Kijas, James
; APPLICANT: Galvert, Sophie
; APPLICANT: Wigh-trowaldh, Gunilla
; APPLICANT: Hedhammar, Ake
; TITLE OF INVENTION: IDENTIFICATION OF CANINE LEUKOCYTE ADHESION DEFICIENCY
; TITLE OF INVENTION: IN DOGS
; FILE REFERENCE: 201515/1001
; CURRENT APPLICATION NUMBER: US/09/393,554
; CURRENT FILING DATE: 1999-09-10
; EARLIER APPLICATION NUMBER: 60/136,099
; EARLIER FILING DATE: 1999-05-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Bos sp.
US-09-393-554-2

Query Match          36.4%; Score 39; DB 4; Length 344;
Best Local Similarity 32.1%; Pred. No. 57;
Matches 9; Conservative 3; Mismatches 4; Indels 12; Gaps 1;

OY 4 ECKRPY-----KELREVGKOR 19
    | : | : | : | : | : |
    Db 197 ECKPPAFRHVLTLDNSKOPETEVEGKOR 224

RESULT 16
5223391-5
; Patent No. 5223391
; APPLICANT: COEN, DONALD M.; DIGARD, PAUL E.
; TITLE OF INVENTION: INHIBITORS OF HERPES SIMPLEX VIRUS
; REPLICATION
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/482,634
; FILING DATE: 21-FEB-1990
; SEQ ID NO: 5
; LENGTH: 520
5223391-5

Query Match          36.4%; Score 39; DB 6; Length 520;
Best Local Similarity 45.0%; Pred. No. 90;
Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

OY 1 SHIECKRPYKELREVGKOR 20
    | : | | | | | : | : | : |
    Db 295 AHEAGKDY--LEIEVGGR 312

RESULT 17
US-08-680-326-36
; Sequence 36, Application US/08680326

```

```

; Patent No. 5925733
; GENERAL INFORMATION:
; APPLICANT: ROSE, TIMOTHY M.
; APPLICANT: BOSCH, MARINX
; APPLICANT: STRAND, KURT
; APPLICANT: TODARO, GEORGE J.
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; TITLE OF INVENTION: FIBROMATOSIS
; NUMBER OF SEQUENCES: 152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,326
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-680-326-36

Query Match          36.4%; Score 39; DB 2; Length 1235;
Best Local Similarity 45.0%; Pred. No. 2,3e+02;
Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

OY 1 SHIECKRPYKELREVGKOR 20
    | : | | | | | : | : | : |
    Db 744 AHEAGKDY--LEIEVGGR 761

RESULT 18
US-08-015-770B-26
; Sequence 26, Application US/08015770B
; Patent No. 5683695
; GENERAL INFORMATION:
; APPLICANT: Shen, De Fen
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Production of recombinant proteins
; TITLE OF INVENTION: containing multiple antigenic determinants linked by
; TITLE OF INVENTION: flexible domains
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: United Biomedical, Inc.
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version  
SOFTWARE: #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,486  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/192,060  
FILING DATE: February 4, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06363/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 542 5070  
TELEFAX: 617 542 8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4544 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-486-52

Query Match 38.3%; Score 41; DB 1; Length 4544;  
Best Local Similarity 56.2%; Pred. No. 4.5e+02;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 CKPKYKELRLEVYKGR 20  
1:11 11 1 11 1  
Db 519 CKPHEHLEFLVYKGR 534

RESULT 13  
US-08-469-658-52  
Sequence 52, Application US/08469658  
Patent No. 5917018  
GENERAL INFORMATION:  
APPLICANT: Th egersen, Hans Christian  
APPLICANT: Holset, Thor Las  
APPLICANT: Ezerodt, Michael  
TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version  
SOFTWARE: #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,658  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/192,060  
FILING DATE: February 4, 1994

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06363/002002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 542 5070  
TELEFAX: 617 542 8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4544 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-658-52

Query Match 38.3%; Score 41; DB 2; Length 4544;  
Best Local Similarity 56.2%; Pred. No. 4.5e+02;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 CKPKYKELRLEVYKGR 20  
1:11 11 1 11 1  
Db 519 CKPHEHLEFLVYKGR 534

RESULT 14  
US-08-406-347A-13  
Sequence 13, Application US/08406347A  
Patent No. 5714374  
GENERAL INFORMATION:  
APPLICANT: ARNOLD, Edward V.  
APPLICANT: ARNOLD, Gall F.  
TITLE OF INVENTION: CHIMERIC RHINOVIROSES  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.50 inch floppy disk, 1.44 mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/406,347A  
FILING DATE: 17-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/304,635  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/041,790  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/583,335  
ATTORNEY/AGENT INFORMATION:  
NAME: Toftennett, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 1984/46203  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-406-347A-13

APPLICANT: Albrecht, Urs  
APPLICANT: Eichele, Gregor  
TITLE OF INVENTION: Mammalian Circadian Rhythm-Like Gene  
FILE REFERENCE: D6039  
CURRENT APPLICATION NUMBER: US/09/150,460B  
CURRENT FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: US 60/058,256  
PRIOR FILING DATE: 1997-09-09  
NUMBER OF SEQ ID NOS: 21  
SEQ ID NO 11  
LENGTH: 1127  
TYPE: PRT  
ORGANISM: D. melanogaster Period  
FEATURE:  
OTHER INFORMATION: Predicted protein sequence compared with RIGUI 4.7  
US-09-150-460B-11

Query Match 40.2%; Score 43; DB 4; Length 1127;  
Best Local Similarity 77.8%; Pred. No. 44;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 ECRKPYKL 12  
DB 392 ECSRPYREL 400

RESULT 10  
US-08-718-388-9  
Sequence 9, Application US/08718388  
Patent No. 6271362  
GENERAL INFORMATION:  
APPLICANT: MORIKAWA, MINORU  
APPLICANT: HARADA, NAOKI  
TITLE OF INVENTION: GENE ENCODING IGG FC REGION-BINDING  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,388  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 0230-111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5405 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-718-388-9

Query Match 39.7%; Score 42.5; DB 4; Length 5405;  
Best Local Similarity 39.1%; Pred. No. 3e+02;  
Matches 9; Conservative 5; Mismatches 4; Indels 5; Gaps 1;

OY 2 HIECRK-----PYKELRLVGVKQ 19  
DB 4824 HVTQEGACAGPHECHLELDGVQ 4846

RESULT 11  
PCT-US91-08177-3  
Sequence 3, Application PC/TUS9108177  
GENERAL INFORMATION:  
APPLICANT: Samal, Siba K  
TITLE OF INVENTION: Bovine Respiratory Syncytial Virus Genes  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti  
STREET: 1201 New York Avenue N.W., suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08177  
FILING DATE: 19911104  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/608,937  
FILING DATE: 05-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Higbet, David W  
REGISTRATION NUMBER: 30,265  
REFERENCE/DOCKET NUMBER: 20509-96711  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4854  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 391 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08177-3

Query Match 38.3%; Score 41; DB 5; Length 391;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 IECKRPYKELRLVGVK 18  
DB 129 IESKSTYKMLKEGE 144

RESULT 12  
US-08-469-486-52  
Sequence 52, Application US/08469486  
Patent No. 5739281  
GENERAL INFORMATION:  
APPLICANT: Thøgersen, Hans Christian  
APPLICANT: Holte, Thor Las  
APPLICANT: Elzerodt, Michael  
TITLE OF INVENTION: Improved method for the refolding of  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts

;; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
;; FILE REFERENCE: 20264262051  
;; CURRENT APPLICATION NUMBER: US/08/397,232A  
;; CURRENT FILING DATE: 1995-04-17  
;; EARLIER APPLICATION NUMBER: 07/947,338  
;; EARLIER FILING DATE: 1992-09-18  
;; EARLIER APPLICATION NUMBER: PCT/US93/08610  
;; EARLIER FILING DATE: 1993-09-17  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 2227  
;; TYPE: PRT  
;; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVQKOR 20  
|||||  
DB 792 SHIECRKPKYKELRLEVQKOR 811

RESULT 7  
US-09-171-387-2  
; Sequence 2, Application US/09171387  
; Patent No. 6280734  
; GENERAL INFORMATION:  
; APPLICANT: RAYCHAUDHURI, GOPA;  
; EMERSON, SUZANNE, U.;;  
; PURCELL, ROBERT, H.;;  
; TITLE OF INVENTION: SIMIAN-HUMAN HAV  
; HAVING A CHIMERIC 2C PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MICROSOFT WORD 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,387  
; FILING DATE: 24-Mar-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/06506  
; FILING DATE: 18-APR-1997  
; APPLICATION NUMBER: US60/015,642  
; FILING DATE: 19-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: William S. Feller  
; REGISTRATION NUMBER: 26,728  
; REFERENCE/DOCKET NUMBER: 2026-4229051  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 2  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2227 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVQKOR 20  
|||||  
DB 792 SHIECRKPKYKELRLEVQKOR 811

RESULT 8  
US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135  
; GENERAL INFORMATION:  
; APPLICANT: NAINAN, OMANA V.  
; APPLICANT: MARIGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DABY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELER: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 79.0%; Score 84.5; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 2.5e-06;  
Matches 18; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 SHIECRKPKYKELRLEVQKOR 20  
|||||  
DB 792 SHIE-KPKYKELRLEVQKOR 810

RESULT 9  
US-09-150-460B-11  
; Sequence 11, Application US/09150460B  
; Patent No. 6190882  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Cheng-Chi

```
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2
```

```
Query Match          100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 SHIECRKPYKELRLEVKGOR 20
    |||||||
Db 792 SHIECRKPYKELRLEVKGOR 811
```

```
RESULT 3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4
```

```
Query Match          100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 SHIECRKPYKELRLEVKGOR 20
    |||||||
Db 792 SHIECRKPYKELRLEVKGOR 811
```

```
RESULT 4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
```

```
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6
```

```
Query Match          100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 SHIECRKPYKELRLEVKGOR 20
    |||||||
Db 792 SHIECRKPYKELRLEVKGOR 811
```

```
RESULT 5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2
```

```
Query Match          100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 SHIECRKPYKELRLEVKGOR 20
    |||||||
Db 792 SHIECRKPYKELRLEVKGOR 811
```

```
RESULT 6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2002, 21:41:06 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432A-39

Perfect score: 107

Sequence: 1 SHIECRKPYKELRLEVGR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	1091	6 US-08-475-886-2	Patent No. 5516630
2	107	100.0	2227	3 US-08-475-886-4	Sequence 2, Appli
3	107	100.0	2227	3 US-08-475-886-6	Sequence 4, Appli
4	107	100.0	2227	3 US-08-475-886-6	Sequence 6, Appli
5	107	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
6	107	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
7	107	100.0	2227	4 US-09-171-387-2	Sequence 2, Appli
8	84.5	79.0	839	1 US-08-087-016-2	Sequence 2, Appli
9	40.2	40.2	1127	4 US-09-150-460B-11	Sequence 11, Appli
10	42.3	39.7	5405	4 US-08-718-368-9	Sequence 9, Appli
11	41	38.3	391	5 PCT-US91-08177-3	Sequence 3, Appli
12	41	38.3	4544	1 US-08-469-486-52	Sequence 52, Appli
13	38.3	38.3	4544	1 US-08-469-486-52	Sequence 52, Appli
14	39	36.4	24	1 US-08-406-347A-13	Sequence 13, Appli
15	39	36.4	344	4 US-09-393-554-2	Sequence 2, Appli
16	39	36.4	520	6 5223391-5	Patent No. 5223391
17	39	36.4	1235	2 US-08-680-326-36	Sequence 36, Appli
18	38	35.5	31	1 US-08-015-770B-26	Sequence 26, Appli
19	38	35.5	31	1 US-08-488-252-15	Sequence 15, Appli
20	38	35.5	102	2 US-08-651-940-3	Sequence 3, Appli
21	38	35.5	123	4 US-09-357-251-22	Sequence 22, Appli
22	38	35.5	155	2 US-08-651-940-4	Sequence 4, Appli
23	38	35.5	256	2 US-08-719-758-2	Sequence 2, Appli
24	38	35.5	256	4 US-09-119-827-2	Sequence 2, Appli
25	38	35.5	364	2 US-08-651-940-2	Sequence 2, Appli
26	38	35.5	364	4 US-09-025-580-37	Sequence 37, Appli
27	38	35.5	364	4 US-09-047-288-2	Sequence 2, Appli

28	38	35.5	372	4	US-09-286-904-24	Sequence 24, Appli
29	37.5	35.0	309	4	US-09-166-412-2	Sequence 2, Appli
30	37.5	35.0	309	4	US-09-347-803-26	Sequence 26, Appli
31	37	34.6	67	6	5196333-10	Patent No. 5196333
32	37	34.6	68	1	US-08-606-789-8	Sequence 8, Appli
33	37	34.6	68	2	US-09-111-348-8	Sequence 8, Appli
34	37	34.6	69	5	PCT-US95-06406A-6	Sequence 6, Appli
35	37	34.6	469	2	US-08-968-751-2	Sequence 2, Appli
36	37	34.6	493	6	US-09-052-089A-1	Sequence 1, Appli
37	37	34.6	5196333-4		Patent No. 5196333	
38	37	34.6	564	4	US-09-360-197-16	Sequence 16, Appli
39	37	34.6	753	4	US-07-861-458C-98	Sequence 98, Appli
40	37	34.6	753	4	US-07-861-458C-99	Sequence 99, Appli
41	37	34.6	855	2	US-08-468-558-2	Sequence 2, Appli
42	37	34.6	855	4	US-08-676-444-2	Sequence 2, Appli
43	37	34.6	972	3	US-08-335-844A-24	Sequence 24, Appli
44	37	34.6	1164	4	US-09-457-708-2	Sequence 2, Appli
45	36.5	34.1	188	2	US-08-933-750C-8	Sequence 8, Appli
46	36.5	34.1	188	4	US-09-234-613-8	Sequence 8, Appli
47	36.5	34.1	313	4	US-09-347-803-25	Sequence 25, Appli
48	36.5	34.1	609	2	US-08-716-301-4	Sequence 4, Appli
49	36	33.6	102	3	US-08-906-616-125	Sequence 125, App
50	36	33.6	102	3	US-08-906-616-125	Sequence 125, App
51	36	33.6	102	4	US-08-639-075A-125	Sequence 125, App
52	36	33.6	102	4	US-09-012-431-125	Sequence 125, App
53	36	33.6	102	4	US-09-012-692-125	Sequence 125, App
54	36	33.6	102	4	US-08-906-613-125	Sequence 125, App
55	36	33.6	183	1	US-07-841-646-23	Sequence 23, Appli
56	36	33.6	183	1	US-08-147-023-23	Sequence 23, Appli
57	36	33.6	183	1	US-08-447-570-23	Sequence 23, Appli
58	36	33.6	183	2	US-08-449-700-23	Sequence 23, Appli
59	36	33.6	183	2	US-08-449-699A-23	Sequence 23, Appli
60	36	33.6	351	3	US-08-861-747-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
5516630-2  
Patent No. 5516630  
APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,  
STEPHEN M.; PORCELL, ROBERT H.; RACANTIELLO, VINCENT R.;  
BAROUDY, BAHIGE M.  
TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/788,262  
FILING DATE: 06-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 256,135  
FILING DATE: 06-OCT-1988  
APPLICATION NUMBER: 654,942  
FILING DATE: 27-SEP-1984  
APPLICATION NUMBER: 537,911  
FILING DATE: 30-SEP-1983  
SEQ ID NO: 2:  
LENGTH: 1091  
5516630-2

Query Match 100.0%; Score 107; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 4.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGR 20  
DB 1029 SHIECRKPYKELRLEVGR 1048

RESULT 2  
US-08-475-886-2  
Sequence 2, Application US/0847586A

**THIS PAGE BLANK (USPTO)**





RESULT 20  
P00431  
genome polypeptide - human hepatitis A virus (strain No. 4) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00431  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 91.2%; Score 93; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 4.9e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 PYKRLRLEVGKORLKYAOE 20  
Db 36 PYKRLRLEVGKORLKYAOE 55

RESULT 21  
JH0135  
genome polypeptide - human hepatitis A virus (strain MS-1) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Apr-1994  
R:Robertson, B.H.; Brown, V.K.; Bradley, D.W.  
Virus Res. 8, 309-316, 1987  
A:Title: Nucleic acid sequence of the VP1 region of attenuated MS-1 hepatitis A virus.  
A:Reference number: JH0135; MUID:88129044  
A:Accession: JH0135  
A:Molecule type: genomic RNA  
A:Residues: 1-319 <ROB>  
A:Cross-references: GB:M22821  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; glycoprotein; polypeptide  
F:1-300/Product: coat protein 1D #status predicted <CPD>  
F:301-319/Product: core protein 2A (fragment) #status predicted <C2A>  
F:237/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.8%; Score 62; DB 2; Length 319;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 PYKRLRLEVGKQ 12  
Db 308 PYKRLRLEVGKQ 319

RESULT 22  
T45997  
hypothetical protein F9D24.280 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000  
C:Accession: T45997  
R:D'Angelis, M.; Vezzi, A.; Modesto, D.; Pigazzini, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23011  
A:Accession: T45997  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <DNA>

A:Cross-references: EMBL:AL137081  
A:Experimental source: cultivar Columbia; BAC clone F9D24  
C:Genetics:  
A:Map position: 3  
A:Insertions: 85/3  
A:Note: F9D24.280  
C:Superfamily: Arabidopsis thaliana hypothetical protein F9D24.210

Query Match 45.1%; Score 46; DB 2; Length 219;  
Best Local Similarity 50.0%; Pred. No. 7.8;  
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
OY 3 KEKRLRLEVGKORLKYAOE 20  
Db 167 KKKKVEGKAKLQRAEE 184

RESULT 23  
T29804  
hypothetical protein C06E4.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T29804  
R:Du, Z.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C06E4.  
A:Reference number: Z20688  
A:Accession: T29804  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-285 <DUZ>  
A:Cross-references: EMBL:U41277; PIDN:AAA82479.1; CESP:C06E4.8  
C:Genetics:  
A:Gene: CESP:C06E4.8  
A:Insertions: 31/2; 127/3; 153/3; 212/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C06E4.8

Query Match 45.1%; Score 46; DB 2; Length 285;  
Best Local Similarity 60.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
OY 4 ELRLRLEVGKORLKYAO 18  
Db 42 ELRLRLEVGKORLKYAO 56

RESULT 24  
D72121  
ATP synthase chain D - Chlamydomonas reinhardtii (strain CWL029)  
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: D72121  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lamm, C.; Fan, J.; Olinger, L.; Grimwood,  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72121  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-209 <ARN>  
A:Cross-references: GB:AE001594; GB:AE001363; MUID:94376341; PIDN:AA018243.1; PID:9437  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: atpD  
C:Superfamily: H+-transporting ATPase chain D

Query Match 44.1%; Score 45; DB 2; Length 209;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

RESULT 16  
GNNTSA  
genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tsarev, S.A.  
Submitted to JIPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:9222597; PIDN:BA00766.1; PID:9222598  
R:Tsarev, S.A.; Emerson, S.O.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and  
A:Reference number: JQ1080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhik  
Submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL>  
A:Cross-references: EMBL:X15461; NID:961971; PIDN:CA03490.1; PID:930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhik  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1499/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 93.1%; Score 95; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 1.4e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGRKRLKYAOEE 20  
|||||:|||||:|||||:|  
Db 803 PYKELRLEVGRKRLKYAOEE 822

RESULT 17  
PQ0435  
genome polyprotein - human hepatitis A virus (strain CY145) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0435  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0435  
A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A:Note: this protein is from the VP1/2A Junction region

C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 92.2%; Score 94; DB 2; Length 55;  
Best Local Similarity 90.0%; Pred. No. 3.3e-08;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGRKRLKYAOEE 20  
|||||:|||||:|||||:|  
Db 35 PYKELRLEVGRKRLKYAOEE 54

RESULT 18  
PQ0437  
genome polyprotein - human hepatitis A virus (strain JM55) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0437  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0437  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 92.2%; Score 94; DB 2; Length 56;  
Best Local Similarity 90.0%; Pred. No. 3.4e-08;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGRKRLKYAOEE 20  
|||||:|||||:|||||:|  
Db 36 PYKELRLEVGRKRLKYAOEE 55

RESULT 19  
GNNTS2  
genome polyprotein - simian hepatitis A virus (strain CY-145) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat prot  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (cynomolgus macaque)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: JQ1180  
R:Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus  
A:Reference number: JQ1180; MUID:91311421  
A:Accession: JQ1180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NAI>  
A:Cross-references: GB:M59286; NID:9329599; PIDN:AAA45473.1; PID:9555083  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.2%; Score 94; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 6.7e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGRKRLKYAOEE 20  
|||||:|||||:|||||:|  
Db 798 PYKELRLEVGRKRLKYAOEE 817

A>Note: Submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORUKYAQEE 20  
Db 799 PYKELRLEVGKORUKYAQEE 818  
|||||

RESULT 12  
GNNVHB  
genome polypeptide - human hepatitis A virus (strain MB9)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
Vpg: protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A>Note: Host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J030303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Missel, T.; Klein, R.; Wimmer, E.; Deinhardt,  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolat  
A:Reference number: J030303; M01D:88045071  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; hyd  
F:1-23/Product: coat protein 1A #status predicted <VP1>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1456/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein Vpg #status predicted <VPG>  
F:1520-1716/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORUKYAQEE 20  
Db 799 PYKELRLEVGKORUKYAQEE 818  
|||||

RESULT 13  
genome polypeptide - human hepatitis A virus (strain EP-35.730) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: P00427  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wilder  
J. Gen. Virol. 73, 1365-1377, 1992

A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; M01D:92300330  
A:Accession: P00427  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
C:Comment: This protein is from the VP1/2A Junction region.  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 98.0%; Score 100; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 3.6e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORUKYAQEE 20  
Db 36 PYKELRLEVGKORUKYAQEE 55  
|||||

RESULT 14  
P00429  
genome polypeptide - human hepatitis A virus (strain PRC16) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: P00429  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; M01D:92300330  
A:Accession: P00429  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
C:Note: This protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 97.1%; Score 99; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 5.3e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORUKYAQEE 20  
Db 36 PYKELRLEVGKORUKYAQEE 55  
|||||

RESULT 15  
P00436  
genome polypeptide - human hepatitis A virus (strain AGM27) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: P00436  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; M01D:92300330  
A:Accession: P00436  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
C:Note: This protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 93.1%; Score 95; DB 2; Length 56;  
Best Local Similarity 90.0%; Pred. No. 2.3e-08;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORUKYAQEE 20  
Db 36 PYKELRLEVGKORUKYAQEE 55  
|||||

F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 102; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 3.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAOEE 20  
DB 799 PYKELRLVGVKQRLKYAOEE 818

RESULT 8  
A03905  
genome polyprotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein 2C; protein 3D  
C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Matzel Jr., J.V.; Purcell, R.H.; Felnstor Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905; MUID:85166289  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polypeptide  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 102; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAOEE 20  
DB 799 PYKELRLVGVKQRLKYAOEE 818

RESULT 9  
GNMYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; core protein 2B; core protein 2C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.; J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with other  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
C:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F:1-23/Product: coat protein 1A #status predicted <VP1>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP4>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>

F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAOEE 20  
DB 799 PYKELRLVGVKQRLKYAOEE 818

RESULT 10  
GNMYHM  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Matjarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MUID:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAJ>  
C:Cross-references: GB:K02990; NID:g329596; PIDN:AAA45472.1; PID:g329597  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAOEE 20  
DB 799 PYKELRLVGVKQRLKYAOEE 818

RESULT 11  
GNMYHM  
genome polyprotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Felnstone, S.M.; Purcell Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison  
A:Reference number: A94149; MUID:87115701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
C:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA45471.1; PID:g329595

A:Accession: PQ0434  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A>Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 36 PYKELRLEVGKORLKYAOEE 55

## RESULT 3

genome polypeptide - human hepatitis A virus (strain TKM002) (fragment)

C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0428

R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992

A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr

A:Reference number: PQ0427; MUID:92300330

A:Accession: PQ0428

A:Molecule type: mRNA

A:Residues: 1-56 <ROB>

A>Note: this protein is from the VP1/2A Junction region

C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 36 PYKELRLEVGKORLKYAOEE 55

## RESULT 4

genome polypeptide - human hepatitis A virus (strain CF-53) (fragment)

C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0432

R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992

A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr

A:Reference number: PQ0427; MUID:92300330

A:Accession: PQ0432

A:Molecule type: mRNA

A:Residues: 1-56 <ROB>

A>Note: this protein is from the VP1/2A Junction region

C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 36 PYKELRLEVGKORLKYAOEE 55

## RESULT 5

PQ0430

genome polypeptide - human hepatitis A virus (strain S23-1) (fragment)

C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0430

R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992

A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge

A:Reference number: PQ0427; MUID:92300330

A:Accession: PQ0430

A:Molecule type: mRNA

A:Residues: 1-56 <ROB>

A>Note: this protein is from the VP1/2A Junction region

C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 36 PYKELRLEVGKORLKYAOEE 55

## RESULT 6

genome polypeptide - human hepatitis A virus (strain LCDC-1) (fragment)

C:Species: human hepatitis A virus  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137

R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R. Nucleic Acids Res. 17, 3594, 1989

A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A vir

A:Reference number: S04137; MUID:89263805

A:Accession: S04137

A:Molecule type: mRNA

A:Residues: 1-341 <AND>

A:Cross-references: EMBL:X14666; NID:g62301; PIDN:CAA32794.1; PID:g4377576

C:Genetics:

A:Gene: VP1

C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; polypeptide

F;2-340/Product: coat protein 1D (VP1) #status predicted <MAT>

Query Match 100.0%; Score 102; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PYKELRLEVGKORLKYAOEE 20  
|||||  
DB 303 PYKELRLEVGKORLKYAOEE 322

## RESULT 7

genome polypeptide - human hepatitis A virus (strain CR326) (fragment)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
C>Note: host Homo sapiens (man)  
C>Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904

R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, J. Virol. 54, 247-255, 1985

A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.

A:Reference number: A03904; MUID:85185648

A:Accession: A03904

A:Molecule type: genomic RNA

A:Residues: 1-852 <LIN>

A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593

C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:15 ; Search time 108.75 Seconds  
(without alignments)  
17,672 Million cell updates/sec

Title: US-09-171-432a-40  
Perfect score: 102  
Sequence: 1 PYKELRLEVGRKRYAOEE 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : PIR.71:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	100.0	55	2 P00433	genome polypeptin
2	102	100.0	56	2 P00434	genome polypeptin
3	102	100.0	56	2 P00428	genome polypeptin
4	102	100.0	56	2 P00432	genome polypeptin
5	102	100.0	56	2 P00430	genome polypeptin
6	102	100.0	341	2 S04137	genome polypeptin
7	102	100.0	852	1 GNNYHA	genome polypeptin
8	102	100.0	1358	2 A03905	genome polypeptin
9	102	100.0	2227	1 GNNYHM	genome polypeptin
10	102	100.0	2227	1 GNNYHR	genome polypeptin
11	102	100.0	2227	1 GNNYHB	genome polypeptin
12	102	100.0	2227	1 GNNYHB	genome polypeptin
13	100	98.0	56	2 P00427	genome polypeptin
14	99	97.1	56	2 P00429	genome polypeptin
15	95	93.1	56	2 P00436	genome polypeptin
16	95	93.1	2230	1 GNNYSA	genome polypeptin
17	94	92.2	55	2 P00435	genome polypeptin
18	94	92.2	56	2 P00437	genome polypeptin
19	94	92.2	839	1 GNNYS2	genome polypeptin
20	93	91.2	56	2 P00431	genome polypeptin
21	62	60.8	319	2 JH0135	genome polypeptin
22	46	45.1	219	2 T45997	hypothetical prote
23	46	45.1	285	2 T29804	hypothetical prote
24	45	44.1	209	2 D72121	hypothetical prote
25	45	44.1	179	2 B86502	ATP synthase chain
26	44	43.1	209	2 A10819	probable exported
27	44	43.1	290	2 C71701	branched-chain ami
28	44	43.1	776	2 T09938	hypothetical prote
29	43	42.2	207	2 H85609	hypothetical prote

30	43	42.2	207	2 A99801	hypothetical prote
31	43	42.2	208	1 F70485	urecill phosphoribo
32	43	42.2	465	2 T25441	hypothetical prote
33	43	42.2	509	2 G81929	probable iron-upta
34	43	42.2	534	2 JC5096	transposase - fung
35	43	42.2	3587	2 T31075	tyrocidine synthet
36	42.5	41.7	179	2 H91049	probable membrane
37	42.5	41.7	179	2 A65027	hypothetical prote
38	42.5	41.7	179	2 E85894	probable membrane
39	42.5	41.7	722	2 AD2222	ATP-dependent heli
40	42	41.2	290	2 B97774	hypothetical prote
41	42	41.2	365	2 F86413	probable cysteine
42	42	41.2	452	2 B83623	probable porin PA0
43	42	41.2	695	2 AH2647	flagellar biosynth
44	42	41.2	723	2 G97429	flagellar biosynth
45	42	41.2	736	2 B82944	ribose/galactose A
46	42	41.2	752	2 A65089	DNA topoisomerase
47	42	41.2	752	2 A45582	DNA topoisomerase
48	42	41.2	752	2 G91116	DNA topoisomerase
49	42	41.2	752	2 G85961	DNA topoisomerase
50	42	41.2	752	2 AF0888	topoisomerase IV c
51	42	41.2	826	2 H84683	hypothetical prote
52	42	41.2	939	2 H71532	valine--tRNA ligas
53	42	41.2	990	2 A86215	protein T6D22.8 (I
54	42	41.2	4131	2 T21085	hypothetical prote
55	41	40.2	68	2 B42243	GTP-binding regula
56	41	40.2	68	2 I39158	conserved hypotet
57	41	40.2	151	2 A72409	hypothetical prote
58	41	40.2	384	2 E84188	aminopeptidase II
59	41	40.2	412	2 E70108	probable alph prot
60	41	40.2	446	2 G70774	

## ALIGNMENTS

RESULT 1  
P00433  
genome polypeptin - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 07-May-1999  
C:Accession: P00433  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nairan, O.V.; Siegl, G.; Wl  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00433  
A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A>Note: This protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptin  
C:Keywords: coat protein; core protein; polypeptin

Query Match 100.0%; Score 102; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 1,7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGRKRYAOEE 20  
DB 35 PYKELRLEVGRKRYAOEE 54

RESULT 2  
P00434  
genome polypeptin - human hepatitis A virus (strain KPH) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 07-May-1999  
C:Accession: P00434  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nairan, O.V.; Siegl, G.; Wl  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; MUID:92300330

**THIS PAGE BLANK (USPTO)**

A:Title: RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit to the beta  
 A:Reference number: A27826; MUID:88011299  
 A:Accession: A27826  
 A:Molecule type: DNA  
 A:Residues: 54-1176 <FAL>  
 A:Cross-references: GB:X05709; GB:M29646; NID:g5514651; PIDN:CAA29180.2; PID:g5514652  
 R:Stizler, S.; Oldenburg, I.; Petersen, G.; Baultz, E.K.F.  
 Gene 100, 155-162, 1991  
 A:Title: Analysis of the promoter region of the housekeeping gene DmRP140 by sequence cc  
 A:Reference number: PQ0154; MUID:91276237  
 A:Accession: PQ0154  
 A:Molecule type: mRNA  
 A:Residues: 1-69 <STR>  
 A:Cross-references: GB:M62972; NID:g157263; PIDN:AAA28476.1; PID:g157267  
 A:Experimental source: embryo  
 C:Genetics:  
 A:Gene: DmRP140  
 A:Cross-references: FlyBase:FBgn0003276  
 A:Introns: 8/1  
 C:Superfamily: DNA-directed RNA polymerase 132K polypeptide  
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription; zinc finger

Query Match 46.6%; Score 48; DB 2; Length 1176;  
 Best Local Similarity 56.2%; Pred. NO. 35;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRLKYAOEELSNEVLP 16  
 :|:|:|:| | |:  
 Db 335 KRKYAKEIIQKEMLP 350

## RESULT 25

S65068  
 DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - tomato  
 C:Species: Lycopersicon esculentum (tomato)  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Jun-1999  
 C:Accession: S65068  
 R:Marilow, D.; Symons, R.H.  
 Plant Mol. Biol. 30, 337-342, 1996  
 A:Title: Sequence analysis of the second largest subunit of tomato RNA polymerase II.  
 A:Reference number: S65068; MUID:96178872  
 A:Accession: S65068  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1191 <MAR>  
 A:Cross-references: EMBL:U28403; NID:g1049067; PIDN:AMC49273.1; PID:g1049068  
 C:Genetics:  
 A:Gene: RPB2  
 C:Superfamily: DNA-directed RNA polymerase 132K polypeptide  
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription

Query Match 46.6%; Score 48; DB 2; Length 1191;  
 Best Local Similarity 56.2%; Pred. NO. 36;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRLKYAOEELSNEVLP 16  
 :|:|:|:| | |:  
 Db 338 KRKYAKEIIQKEMLP 353

Search completed: June 16, 2002, 00:05:17  
 Job time: 791 sec







A:Reference number: JQ1180; MID:91311421  
A:Accession: JQ1180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NAI>  
A:Cross-references: GB:M59286; NID:q329599; PIDN:AAA5473.1; PID:q555083  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; glycoprotein; polypeptide  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-491/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F:261,312,728,756/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 91.3%; Score 94; DB 1; Length 839;  
Best Local Similarity 85.0%; Pred. No. 1.3e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNELPPK 20  
|||:|||||:|||||  
Db 809 QREYAREELSNELPPK 828

RESULT 10  
S7448  
regulatory protein pcr - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein sll1408  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 18-Aug-2000  
C:Accession: S7448  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, D. N. Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A:Reference number: S74322; MID:97061201  
A:Accession: S7448  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-346 <KAN>  
A:Cross-references: EMBL:D50899; GB:AB001339; NID:gl651650; PIDN:BA16600.1; PID:gl65167  
C:Note: the nucleotide sequence was submitted to the EMBL data library, June 1996  
C:Genetics:  
A:Gene: pcr  
C:Superfamily: *Synechocystis* regulatory protein pcr  
C:Keywords: transcription regulation

Query Match 54.4%; Score 56; DB 2; Length 346;  
Best Local Similarity 61.1%; Pred. No. 0.48;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNELPP 18  
|||:|||||:|||||  
Db 241 ERVAYAOEELVKOVLDP 258

RESULT 11  
T43701  
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain [imported] - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T43701  
R:Sidow, A.; Thomas, W. K.  
Curr. Biol. 4, 596-603, 1994  
A:Title: A molecular evolutionary framework for eukaryotic model organisms.  
A:Reference number: Z22636; MID:95041334  
A:Accession: T43701  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-859 <STD>  
A:Cross-references: EMBL:U010333; NID:g520506; PIDN:AAA50224.1; PID:g520507

C:Superfamily: DNA-directed RNA polymerase 132k polypeptide  
C:Keywords: nucleotidyltransferase

Query Match 50.5%; Score 52; DB 2; Length 859;  
Best Local Similarity 62.5%; Pred. No. 5.8;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNELPP 16  
|||:|||||:|||||  
Db 141 QRIKAREILQKELPP 156

RESULT 12  
E88445  
protein C26E6.4 [imported] - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: E88445  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio  
A:Reference number: A75000; MID:99069613; PMID:9851916  
A:Note: See websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A:Accession: E88445  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1193 <STO>  
A:Cross-references: GB:chr\_III; PIDN:AAA21158.1; PID:g532805; GSPDB:GN00021; CESP:C26  
C:Genetics:  
A:Gene: C26E6.4  
A:Map position: 3  
C:Superfamily: DNA-directed RNA polymerase 132k polypeptide

Query Match 50.5%; Score 52; DB 2; Length 1193;  
Best Local Similarity 62.5%; Pred. No. 8.4;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNELPP 16  
|||:|||||:|||||  
Db 340 QRIKAREILQKELPP 355

RESULT 13  
T50995  
related to cytoskeleton assembly control protein SLA1 [imported] - *Neurospora crassa*  
N:Alternate names: protein B7F18.140  
C:Species: *Neurospora crassa*  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50995  
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakutu  
submitted to the Protein Sequence Database, July 2000  
A:Reference number: Z25286  
A:Accession: T50995  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1119 <SCH>  
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.140  
A:Experimental source: BAC clone B7F18; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B7F18.140  
A:Map position: 6  
A:Introns: 66/3; 123/2; 495/1

Query Match 49.5%; Score 51; DB 2; Length 1119;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNELPP 18  
|||:|||||:|||||

C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wessel, T.; Kiehn, R.; Wimmer, E.; Deinhardt, Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolat  
A:Reference number: J50303; MUID:88045071  
A:Accession: J50303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; hydr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: core protein 2A #status predicted <2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein VPg #status predicted <VPg>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYA0EELSNEVLP PPRK 20  
||||| ||||||| ||||||| ||||||| |||||||  
Db 810 ORLKYA0EELSNEVLP PPRK 829

RESULT 7  
GNNYMK  
genome polypeptide - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro  
NA polymerase (EC 2.7.7.48); protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Comen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Felstone, S.M.; Purcell, R  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison wit  
A:Reference number: A94149; MUID:87175701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M16632; NID:G329594; PIDN:AAA5471.1; PID:G329595  
A:Note: Submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 95.1%; Score 98; DB 1; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 9.1e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYA0EELSNEVLP PPRK 20  
||||| ||||||| ||||||| ||||||| |||||||

Db 810 ORLKYA0EELSNEVLP PPRK 829

RESULT 8  
GNNYSA  
genome polypeptide - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
C:Species: simian hepatitis A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tsarev, S.A.  
Submitted to JIPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:DD0924; NID:G222597; PIDN:BAA00766.1; PID:G222598  
R:Tsarev, S.A.; Emerson, S.O.; Balaian, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure  
A:Reference number: J01080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balaian, M.S.; Kusov, Y.Y.; Andjapardze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chlzh  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:G61971; PIDN:CAA33490.1; PID:G930268  
R:Balaian, M.S.; Kusov, Y.Y.; Andjapardze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chlzh  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-245/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-994/Product: core protein 2A #status predicted <C2A>  
F:995-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 94.2%; Score 97; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 1.3e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ORLKYA0EELSNEVLP PPRK 20  
||||| ||||||| ||||||| ||||||| |||||||  
Db 814 ORLKYA0EELSNEVLP PPRK 833

RESULT 9  
GNNYS2  
genome polypeptide - simian hepatitis A virus (strain CY-145) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat prot  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (cynomolgus macaque)  
C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: J01180  
R:Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balaian, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus

C:Accession: A03904  
R:Jinney, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.W.  
J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904, MIMD:85185648  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <LIN>  
A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 103; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLP PPK 20  
|||||

DB 810 ORLKYAOEELSNEVLP PPK 829

RESULT 3  
A03905  
genome polypeptide (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein  
C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
A:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Melele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinstone  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905, MIMD:85166289  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 103; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 8.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLP PPK 20  
|||||

DB 810 ORLKYAOEELSNEVLP PPK 829

RESULT 4  
GNNYHM  
genome polypeptide - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
A:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with d  
A:Reference number: A25981, MIMD:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA

A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-223/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLP PPK 20  
|||||

DB 810 ORLKYAOEELSNEVLP PPK 829

RESULT 5  
GNNYHR  
genome polypeptide - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
A:Accession: A03903  
R:Marjarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MIMD:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAJ>  
A:Cross-references: GB:K02990; NID:g329596; PIDN:AAA45472.1; PID:g329597  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVLP PPK 20  
|||||

DB 810 ORLKYAOEELSNEVLP PPK 829

RESULT 6  
GNNYHB  
genome polypeptide - human hepatitis A virus (strain MB9)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
Vg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:16 ; Search time 108.75 Seconds

(without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-4332a-41

Perfect score: 103

Sequence: 1 ORLKYAOEELSNEVLPPRRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : PIR.71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	341	2	S04137
2	103	100.0	852	2	GNNYHA
3	103	100.0	1358	2	A03905
4	103	100.0	2227	1	GNNYHM
5	103	100.0	2227	1	GNNYHR
6	103	100.0	2227	1	GNNYHB
7	98	95.1	2227	1	GNNYMK
8	97	94.2	2230	1	GNNYSA
9	94	91.3	839	1	GNNYS2
10	56	54.4	346	2	S74448
11	52	50.5	859	2	T43701
12	52	50.5	1193	2	E88445
13	51	49.5	1119	2	T50995
14	49.5	48.1	443	2	E82046
15	49	47.6	55	2	P00433
16	49	47.6	56	2	P00434
17	49	47.6	56	2	P00428
18	49	47.6	56	2	P00427
19	49	47.6	56	2	P00429
20	49	47.6	56	2	P00432
21	49	47.6	56	2	P00430
22	49	47.6	1174	2	S28976
23	48	46.6	442	2	A71969
24	48	46.6	1176	2	A27826
25	48	46.6	1191	2	S65068
26	47	45.6	592	1	LIBY
27	47	45.6	6642	2	T29757
28	46	44.7	338	2	I56893
29	46	44.7	480	1	FMPUB3

30	46	44.7	483	2	T01053	hypothetical prote
31	45	43.7	756	2	JC5590	cytochrome oxidase
32	45	43.7	1188	2	T05846	DNA-directed RNA p
33	45	43.7	2420	2	A84652	hypothetical prote
34	44.5	43.2	1609	2	T01797	hypothetical prote
35	44	42.7	56	2	P00436	genome polypeptide
36	44	42.7	223	2	P00514	hemagglutinin HA2
37	44	42.7	303	2	S61723	phosphoglycerate m
38	44	42.7	443	2	AH0013	ATP-binding heat s
39	44	42.7	510	2	S42705	probable phosphogl
40	44	42.7	569	2	T43531	probable potassium
41	44	42.7	694	2	T10565	hypothetical prote
42	44	42.7	914	2	S48333	ORC1 protein - yea
43	43.5	42.2	3678	2	G82530	hypothetical prote
44	43.5	42.2	3678	2	S28916	dystrophin - mouse
45	43.5	42.2	3685	1	A27605	dystrophin, muscle
46	43	41.7	36	2	S78721	protein YGL006w-a
47	43	41.7	36	2	P00437	genome polypeptide
48	43	41.7	120	2	D69879	alkaline-shock pro
49	43	41.7	163	2	PC4186	hemagglutinin 2 ch
50	43	41.7	268	2	S48942	hypothetical prote
51	43	41.7	277	2	T11680	hypothetical prote
52	43	41.7	496	2	G64527	hypothetical prote
53	43	41.7	574	1	HMIVBM	hemagglutinin prec
54	43	41.7	576	2	S03300	hemagglutinin prec
55	43	41.7	577	2	A97195	uncharacterized AB
56	43	41.7	578	2	S03299	hemagglutinin prec
57	43	41.7	578	2	S03301	hemagglutinin prec
58	43	41.7	583	1	HMIVBS	hemagglutinin prec
59	43	41.7	583	1	HMIVHO	hemagglutinin prec
60	43	41.7	585	1	HMIVBJ	hemagglutinin prec

## ALIGNMENTS

RESULT 1  
S04137  
genome polypeptide - human hepatitis A virus (strain LCD-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
R:Accession: S04137  
R:Anderson, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A vir  
A:Reference number: S04137; MUID:89263805  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>  
A:Cross-references: EMBL:X14666; NID:962301; PIDN:CNA32794.1; PID:94377576  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; polypeptide  
F:2-340/Product: coat protein ID (VP1) #status predicted <MAY>

Query Match 100.0%; Score 103; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORLKYAOEELSNEVLPPRRK 20  
|||||  
DB 314 ORLKYAOEELSNEVLPPRRK 333

RESULT 2  
GNNYHA  
genome polypeptide - human hepatitis A virus (strain CR326) (fragment)  
M:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999



A:Accession: A54562  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 7'p,1316-1595 <SUN>  
 A:Cross-references: GB:M29866; NID:9203200; PIDN:AAA04087.1; PID:9554423  
 R:Jacobs, J.W.; Rubin, J.S.; Huggl, T.E.; Bogardt, R.A.; Maritz, I.K.; Daniels, J.S.; Dab  
 Biochemstry 17, 5031-5038, 1978  
 A:Title: Purification, characterization, and amino acid sequence of rat anaphylatoxin (C  
 A:Reference number: A01260; MUID:79062262  
 A:Accession: A01260  
 A:Molecule type: protein  
 A:Residues: 671-703, 'K', 705-720, 'K', 723-748 <UAC>  
 A:Note: three disulfide bonds are present  
 R:Suwa, Y.; Kudo, I.; Imaizumi, A.; Okada, M.; Kamimura, T.; Suzuki, Y.; Chang, H.W.; Ha  
 Proc. Natl. Acad. Sci. U.S.A. 87, 2395-2399, 1990  
 A:Title: Proteinaceous inhibitors of phospholipase A-2 purified from inflammatory sites  
 A:Reference number: A35979; MUID:90207203  
 A:Accession: B35979  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 'X', 998-1005 <SUN>  
 A:Accession: A35979  
 A:Molecule type: protein  
 A:Residues: 'X', 961-962, 'P', 964-969 <SU2>  
 R:Nakagawa, H.; Komorita, N.  
 Biochem. Biophys. Res. Commun. 194, 1181-1187, 1993  
 A:Title: Complement component C3-derived neutrophil chemotactic factors purified from ex  
 A:Reference number: PNO566; MUID:93356786  
 A:Accession: PNO567  
 A:Molecule type: protein  
 A:Residues: 568-592 <NAK>  
 A:Note: amino end of a C3-derived peptide designated exudate neutrophil chemotactic fact  
 A:Accession: PNO566  
 A:Molecule type: protein  
 A:Residues: 671-687 <NA2>  
 A:Note: amino end of peptide designated neutrophil chemotactic factor 1 and probably ide  
 R:Kivilanen, P.C.; Capuliong, R.B.; Harkins, R.N.; Desombre, E.R.  
 Blochem. Biophys. Res. Commun. 158, 898-905, 1989  
 A:Title: The estrogen-responsive 110K and 74K rat uterine secretory proteins are structu  
 A:Reference number: A32281; MUID:89149812  
 A:Accession: A32281  
 A:Molecule type: protein  
 A:Residues: 25-41 <KUT>  
 A:Experimental source: 17beta-estradiol-stimulated uterus of immature rat  
 A:Note: the authors treat this 74K uterine secretory protein, identical as far as sequen  
 ent  
 C:Comment: Complement C3 contains two chains, formed by removal of four residues and lit  
 alternate complement pathways, releases the C3a anaphylatoxin from the amino end of t  
 native complement pathway C3/C5 convertase.  
 C:Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
 C:Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa  
 e classical complement pathway C3/C5 convertase. The activity of C3b is regulated by pro  
 C:Comment: The major site of synthesis of this plasma protein is the liver.  
 C:Superfamily: alpha-2-macroglobulin  
 C:Keywords: acute phase; chemotaxis; complement alternate pathway; complement pathway; g  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-666/Product: complement C3 and C3b beta chain #status predicted <C3BA>  
 F:25-666,671-1663/Product: complement C3 #status predicted <C3>  
 F:25-666,749-1663/Product: complement C3b #status predicted <C3B>  
 F:671-1663/Product: complement C3 alpha chain #status predicted <CC3A>  
 F:671-748/Product: C3a anaphylatoxin #status experimental <C3T>  
 F:749-1663/Product: complement C3b alpha' chain #status predicted <C3BA>  
 F:946-1303/Product: C3dk fragment #status predicted <CDK>  
 F:1002-1303/Product: C3d fragment #status predicted <CDK>  
 F:1424-1457/Region: properdin binding  
 F:558-816,626-661,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506-15  
 F:748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted  
 F:939,1617/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:1010-1013/Cross-link: thiolester (Cys-Gln) #status predicted  
 F:1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted  
 F:1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted

Query Match 40.7%; Score 43.5; DB 1; Length 1663;  
 Best Local Similarity 47.4%; Pred. No. 1.7e+02;  
 Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

Oy 1 SHIECRKPYKELRLEVGKO 19  
 Db 1586 SHVKCR---NALKIQKGKO 1601

## RESULT 25

F96568  
 Probable 1lipase, 20450-21648 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence.revision 02-Mar-2001 #text.change 31-Mar-2001  
 C:Accession: F96568  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar,  
 ansen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Frisier, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: F96568  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-532 <STO>  
 A:Cross-references: GB:AE005173; NID:96862937; PIDN:AAF30325.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F14G24.3  
 A:Map position: 1

Query Match 40.2%; Score 43; DB 2; Length 332;  
 Best Local Similarity 54.5%; Pred. No. 40;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 2 HIECRKPYKEL 12  
 Db 129 HVRCSDPYKDL 139

Search completed: June 16, 2002, 00:05:15  
 Job time: 789 sec



GNMYS2  
genome polypeptide - simian hepatitis A virus (strain CY-145) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (Cynomolgus macaque)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: J01180  
R:Naiman, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus mac  
A:Reference number: J01180; MUID:913111421  
A:Accession: J01180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NAI>  
A:Cross-references: GB:M59286; NID:g329599; PIDN:AAA45473.1; PID:g555083  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; glycoprotein; polypeptide  
F:1-23/Product: coat protein 1A #status predicted <VP3>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-491/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 79.0%; Score 84.5; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 2.5e-05;  
Matches 18; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 SHICRKPYLELREVGKQ 20  
|||:|||||:|||||:  
Db 792 SHICRKPYLELREVGKQ 810

RESULT 21  
P00437  
genome polypeptide - human hepatitis A virus (strain JM55) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: P00437  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Naiman, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00437  
A:Molecule type: RNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 75.7%; Score 81; DB 2; Length 56;  
Best Local Similarity 84.2%; Pred. No. 5.6e-06;  
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SHICRKPYLELREVGKQ 19  
|||:|||||:|||||:  
Db 29 SHICRKPYLELREVGKQ 47

RESULT 22  
S62541  
hypothetical protein SPAC12G12.10 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000  
C:Accession: S62541; T37591  
R:Devlin, K.; Odell, C.; Churcher, C.M.  
Submitted to the EMBL Data Library, November 1995  
A:Reference number: S62532  
A:Accession: S62541  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-420 <DEV>

A:Cross-references: EMBL:266568; NID:g1052518; PIDN:CAA91505.1; PID:g1052528  
R:Devlin, K.; Odell, C.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsby, S.V.  
Submitted to the EMBL Data Library, November 1995  
A:Reference number: 221727  
A:Accession: T37591  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-420 <DEV>  
A:Cross-references: EMBL:266568; PIDN:CAA91505.1; GSPDB:GN00066; SPDB:SPAC12G12.10  
A:Experimental source: strain 972h-; cosmid c12G12  
C:Genetics:  
A:Gene: SPAC12G12.10  
A:Map position: 1L

Query Match 43.0%; Score 46; DB 2; Length 420;  
Best Local Similarity 46.7%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 4 ECRKPYLELREVGK 18  
|||:|||||:|||||:  
Db 376 DCSLPFKEMRVDDGK 390

RESULT 23  
C84040  
hypothetical protein BH3123 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: C84040  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: A83650; MUID:20512582; PMID:11056132  
A:Accession: C84040  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-374 <STO>  
A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BA006842.1; GSPDB:G  
C:Experimental source: strain C-125  
A:Genetics:  
A:Gene: BH3123

Query Match 41.1%; Score 44; DB 2; Length 374;  
Best Local Similarity 62.5%; Pred. No. 31;  
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 4 ECRKPYLELREVGKQ 19  
|||:|||||:|||||:  
Db 124 ELQKKEKELRDEVGKE 139

RESULT 24  
C3RT  
complement C3 precursor - rat  
N:Alternate names: 37k phospholipase A2 inhibitory protein  
N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 04-Dec-1992 #sequence\_revision 07-Oct-1994 #text\_change 18-Jun-1999  
C:Accession: S15764; A54562; A01260; B35979; A35979; PNO567; PNO566; A32281; S08692  
R:Mitsun, Y.; Schda, M.; Ikehara, Y.  
Nucleic Acids Res. 18, 2178, 1990  
A:Title: Nucleotide and deduced amino acid sequence of rat complement C3.  
A:Reference number: S15764; MUID:90245672  
A:Accession: S15764  
A:Molecule type: mRNA  
A:Residues: 1-1663 <MIS>  
A:Cross-references: EMBL:X52477; NID:g56953; PIDN:CAA36716.1; PID:g56954  
R:Sundstrom, S.A.; Komm, B.S.; Ponce-de-Leon, H.; Yi, Z.; Teuscher, C.; Lytle, C.R.  
J. Biol. Chem. 264, 16941-16947, 1989  
A:Title: Estrogen regulation of tissue-specific expression of complement C3.  
A:Reference number: A54562; MUID:89380332

OY 1 SHIECRKPKYKELRLEVGRKOR 20  
||||| ||||||| ||||||| |||||||  
Db 29 SHIECRKPKYKELRLEVGRKOR 48

RESULT 16  
PQ0433  
genome polyprotein - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
A:Accession: PQ0433  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wildel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0433  
A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A>Note: this protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 81.3%; Score 87; DB 2; Length 55;  
Best Local Similarity 90.0%; Pred. No. 6.1e-07;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGRKOR 20  
||||| ||||||| ||||||| |||||||  
Db 28 SHIECRKPKYKELRLEVGRKOR 47

RESULT 17  
PQ0436  
genome polyprotein - human hepatitis A virus (strain AGM27) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
A:Accession: PQ0436  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wildel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0436  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A>Note: this protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 80.4%; Score 86; DB 2; Length 56;  
Best Local Similarity 85.0%; Pred. No. 8.9e-07;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGRKOR 20  
||||| ||||||| ||||||| |||||||  
Db 29 SHIECRKPKYKELRLEVGRKOR 48

RESULT 18  
CNVSA  
genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
A:Accession: A30470; S04885; S03965  
R:Tsarev, S.A.  
submitted to JIPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BA00766.1; PID:g222598

R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure  
A:Reference number: J01080; MUID:91311420  
A:Contents: annotation

A>Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizh  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:961971; PIDN:CA03490.1; PID:9930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizh  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965

A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
P:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 80.4%; Score 86; DB 1; Length 2230;  
Best Local Similarity 85.0%; Pred. No. 3.8e-05;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SHIECRKPKYKELRLEVGRKOR 20  
||||| ||||||| ||||||| |||||||  
Db 796 SHIECRKPKYKELRLEVGRKOR 815

RESULT 19  
PQ0435  
genome polyprotein - human hepatitis A virus (strain Cyl45) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
A:Accession: PQ0435  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wl  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0435  
A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A>Note: this protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 79.0%; Score 84.5; DB 2; Length 55;  
Best Local Similarity 90.0%; Pred. No. 1.5e-06;  
Matches 18; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 SHIECRKPKYKELRLEVGRKOR 20  
||||| ||||||| ||||||| |||||||  
Db 29 SHIECRKPKYKELRLEVGRKOR 47

RESULT 20

Query Match  
Best Local Similarity 97.2%; Score 104; DB 2; Length 56;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGKOR 48

RESULT 11  
JH0135  
genome polyprotein - human hepatitis A virus (strain MS-1) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Apr-1994  
C:Accession: JH0135  
R:Robertson, B.H.; Brown, V.K.; Bradley, D.W.  
A>Title: Nucleic acid sequence of the VP1 region of attenuated MS-1 hepatitis A virus.  
A:Reference number: JH0135; MUID:88129044  
A:Accession: JH0135  
A:Molecule type: genomic RNA  
A:Residues: 1-319 <ROB>  
A:Cross-references: GB:M22821  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1300/Product: coat protein 1D (status predicted <CDB>  
F:301-319/Product: core protein 2A (fragment) (status predicted <C2A>  
F:237/Binding site: carbohydrate (Asn) (covalent) (status predicted)

Query Match  
Best Local Similarity 95.3%; Score 102; DB 2; Length 319;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 19  
|||||  
Db 301 SHIECRKPYKELRLEVGKOR 319

RESULT 12  
PQ0430  
genome polyprotein - human hepatitis A virus (strain S23-1) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0430  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0430  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match  
Best Local Similarity 93.5%; Score 100; DB 2; Length 56;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGKOR 48

RESULT 13  
PQ0432  
genome polyprotein - human hepatitis A virus (strain CF-53) (fragment)

C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0432  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0432  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match  
Best Local Similarity 90.7%; Score 97; DB 2; Length 56;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGKOR 48

RESULT 14  
PQ0428  
genome polyprotein - human hepatitis A virus (strain TKM002) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0428  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0428  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match  
Best Local Similarity 89.7%; Score 96; DB 2; Length 56;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 29 SHIECRKPYKELRLEVGKOR 48

RESULT 15  
PQ0434  
genome polyprotein - human hepatitis A virus (strain KPH) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0434  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0434  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match  
Best Local Similarity 88.8%; Score 95; DB 2; Length 56;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match	100.0%;	Score 107;	DB 1;	Length 2227;
Best Local Similarity	100.0%;	Pred. No. 1.7e-08;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY 1 SHIECRKPYKELRLEVQKOR 20  
      |||||  
Db 792 SHIECRKPYKELRLEVQKOR 811
```

7  
GENBANK  
genome polyprotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; core protein 3A; core protein 3B; core protein 3C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 16-Jul-1999  
C:Accession: A94149; A25514; A94508  
R:Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COOH>  
A:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA45471.1; PID:g329595  
A:Note: submitted to Genbank, August 1987  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-401/Product: coat protein 1B #status predicted <P1B>  
F:402-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1676/Product: protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match	100.0%;	Score 107;	DB 1;	Length 2227;
Best Local Similarity	100.0%;	Pred. No. 1.7e-08;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY 1 SHIECRKPYKELRLEVQKOR 20  
    |||||  
Db 792 SHIECRKPYKELRLEVQKOR 811
```

**RESULT**      **8**

GNNYHB  
genome polyprotein - human hepatitis A virus (strain MBB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; core protein 2B; RNA polymerase (EC 2.7.7.48), protein 3D  
Vpg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A>Note: host Homo sapiens (man)  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303

R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Klein, R.; Wimmer, E.; Deinhardt,  
virus Res. 8, 153-171, 1987

A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso  
A:Reference number: J50303; MOID:88045071

A;Accession: JS0303  
A;molecule type: genomic RNA

A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein, core protein, cysteine proteinase; genome-Linked protein; h  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: core protein 2A #status predicted <P2>  
F:981-1108/Product: core protein 2B #status predicted <P2>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-Linked protein Vrg #status predicted <VPG>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match	100.0%;	Score 107;	DB 1;	length 2227;
Best Local Similarity	100.0%;	Pred. No. 1.7e-08;		
Matches	20;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

```
QY 1 SHIECRKPYKELRLEVGVKOR 20
    |||||
Db 792 SHIECRKPYKELRLEVGVKOR 811
```

RESULT 9  
P00427  
genome polyprotein - human hepatitis A virus (strain EP-35.730) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 07-May-1999  
C:Accession: P00427  
R:Robertson, B. H.; Jansen, R. W.; Khanna, B.; Totsuka, A.; Nainan, O. V.; Slegel, G.; Whit  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00427  
A:Molecule type: mRNA  
A:Residues: 1-56 <R0B>  
C:Comment: This protein is from the VP1/2A junction region.  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

```

Query Match      98.1%; Score 105; DB 2; Length 56;
Best Local Similarity 95.0%; Pred. No. 8.3e-10;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 1 SHIECRKPYKELRLEVQKOR 20
    |||||:|||||
Db 29 SHIECRKPYKELRMEVGKOR 48
```

RESULT 10  
P00429 genome polyprotein - human hepatitis A virus (strain PRCL6) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: P00429  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; W  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00429  
A:Molecule type: mRNA  
A:Residues: 1-56 <R08A>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>  
A:Cross-references: EMBL:X14666; NID:g62301; PIDN:CAA32794.1; PID:g4377576  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; polyprotein  
F:2-340/Product: coat protein 1D (VP1) #status predicted <MAT>

Query Match 100.0%; Score 107; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.5e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 296 SHIECRKPYKELRLEVGKOR 315

## RESULT 3

GNMYHA  
genome polyprotein - human hepatitis A virus (strain CR326) (fragment)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
A>Note: host Homo sapiens (man)  
C>Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904  
R:Linneyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitrá, S.M.  
J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904; MUID:85185648  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <LIN>  
A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <CIA>  
F:246-491/Product: coat protein 1B #status predicted <CIB>  
F:492-836/Product: coat protein 1C #status predicted <CIB>  
F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 107; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 6.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 792 SHIECRKPYKELRLEVGKOR 811

## RESULT 4

A03905  
genome polyprotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein  
C:Species: human hepatitis A virus  
C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Matzel Jr., J.V.; Purcell, R.H.; Feinston  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905; MUID:85162689  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <CIA>  
F:246-491/Product: coat protein 1B #status predicted <CIB>  
F:492-836/Product: coat protein 1C #status predicted <CIB>  
F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>

F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 107; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 792 SHIECRKPYKELRLEVGKOR 811

## RESULT 5

GNMYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A>Note: host Homo sapiens (man)  
C>Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SHIECRKPYKELRLEVGKOR 20  
|||||  
Db 792 SHIECRKPYKELRLEVGKOR 811

## RESULT 6

GNMYHR  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A>Note: host Homo sapiens (man)  
C>Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MUID:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAJ>  
A:Cross-references: GB:K02990; NID:g329596; PIDN:AAA45472.1; PID:g329597  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <CIA>

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2002, 23:52:06 ; Search time 108.75 Seconds

(without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-432a-39

Perfect score: 107

Sequence: 1 SHIECRKPYKELRELVGKQR 20

Scoring table: BLOSUMP62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : PIR.71.\*

1: pirl.\*  
2: pirl.\*  
3: pirl.\*  
4: pirl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	56	2 P00431	genome polypotein
2	107	100.0	341	2 S04137	genome polypotein
3	107	100.0	852	1 GNNYHA	genome polypotein
4	107	100.0	1358	2 A03905	genome polypotein
5	107	100.0	2227	1 GNNYHA	genome polypotein
6	107	100.0	2227	1 GNNYHA	genome polypotein
7	107	100.0	2227	1 GNNYHA	genome polypotein
8	107	100.0	2227	1 GNNYHA	genome polypotein
9	105	98.1	56	2 P00427	genome polypotein
10	104	97.2	56	2 P00429	genome polypotein
11	102	95.3	319	2 JH0135	genome polypotein
12	100	93.5	56	2 P00430	genome polypotein
13	97	90.7	56	2 P00432	genome polypotein
14	96	89.7	56	2 P00428	genome polypotein
15	95	88.8	56	2 P00434	genome polypotein
16	87	81.3	55	2 P00433	genome polypotein
17	86	80.4	56	2 P00436	genome polypotein
18	86	80.4	2230	1 GNNYSA	genome polypotein
19	84.5	79.0	55	2 P00435	genome polypotein
20	84.5	79.0	839	1 GNNYSA	genome polypotein
21	81	75.7	56	2 P00437	genome polypotein
22	46	43.0	420	2 S62541	hypothetical prote
23	44	41.1	374	2 C84040	hypothetical prote
24	43.5	40.7	1663	1 C3RT	complement C3 prec
25	43	40.2	332	2 F96568	probable lipase, 2
26	43	40.2	339	2 T18926	hypothetical prote
27	43	40.2	343	2 A11823	30S ribosomal prot
28	43	40.2	346	2 T27896	hypothetical prote
29	43	40.2	391	1 VHN23	nucleocapsid prote

30	43	40.2	554	2 T49917	hypothetical prote
31	42.5	39.7	140	2 E64364	hypothetical prote
32	42	39.3	121	2 T17708	hypothetical prote
33	42	39.3	214	2 E82679	chaperone xrl452
34	42	39.3	286	2 S07532	puff 11/9-1 protei
35	42	39.3	329	2 T25067	hypothetical prote
36	42	39.3	723	2 S62902	legumlin 2 precurs
37	42	39.3	723	2 D71091	methionine--tRNA 1
38	42	39.3	861	2 B84963	DNA topoisomerase
39	42	39.3	1087	2 T16876	hypothetical prote
40	42	39.3	1157	2 T43259	pyruvate (flavodox
41	42	39.3	1270	2 T22615	hypothetical prote
42	41.5	38.8	286	2 H87506	hypothetical prote
43	41.5	38.8	606	2 H97431	phosphogluconate d
44	41.5	38.8	606	2 A12649	phosphogluconate d
45	41	38.3	144	2 H96511	AP2 domain contain
46	41	38.3	232	2 A13559	AMP-dependent prot
47	41	38.3	298	2 B83823	endonuclease IV BH
48	41	38.3	338	2 A86412	protein FIK23.10
49	41	38.3	364	2 C64326	succinyl-CoA synth
50	41	38.3	370	2 H64480	hypothetical prote
51	41	38.3	391	1 VHN21	nucleocapsid prote
52	41	38.3	391	1 VHN2B4	nucleocapsid prote
53	41	38.3	391	1 J01533	nucleocapsid prote
54	41	38.3	467	1 VHN2	nucleocapsid prote
55	41	38.3	635	2 D84920	hypothetical prote
56	41	38.3	640	2 A86657	fructose-bisphosph
57	41	38.3	1549	2 T13940	ankyrin - fruit fl
58	41	38.3	4543	1 A53102	alpha-2-macroglobu
59	41	38.3	4543	1 S02392	alpha-2-macroglobu
60	41	38.3	4545	1 S25111	alpha-2-macroglobu

## ALIGNMENTS

RESULT 1  
P00431  
genome polypotein - human hepatitis A virus (strain No. 4) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993. #text\_change 07-May-1999  
C:Accession: P00431  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Sleight, G.; Wl  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: P00427; MUID:92300330  
A:Accession: P00431  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polypotein  
C:Keywords: coat protein; core protein; polypotein

Query Match 100.0%; Score 107; DB 2; Length 56;  
Best local similarity 100.0%; Pred. No. 4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Or 1 SHIECRKPYKELRELVGKQR 20  
Db 29 SHIECRKPYKELRELVGKQR 48  
|||||

RESULT 2  
S04137  
genome polypotein - human hepatitis A virus (strain LCD-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137  
R:Anderson, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A vir  
A:Reference number: S04137; MUID:89263805

CC THROMBOCYTOPENIA, HAIR LOSS, DIARRHEA, FEVER, MARKED WEIGHT LOSS,  
CC CEREBELLAR ATAXIA, AND NEUROLOGIC SYMPTOMS, PROGRESSING TO  
CC SEMICOMA.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U09178; AAB57474.1; -;  
DR EMBL: U20938; AAB51366.1; -;  
DR EMBL: X95670; CAA64973.1; -;  
DR HSSP: P26886; 1ERD.  
DR MIM: 274270; -;  
DR InterPro: IPR001450; 4Fe4S\_ferredoxin.  
DR InterPro: IPR001295; DHO\_dh.  
DR InterPro: IPR001327; FAD\_pyr\_redox.  
DR InterPro: IPR003009; FMN\_enzyme.  
DR Pfam: PF01180; DHodehase; 1.  
DR Pfam: PF00037; fer4; 2.  
DR Pfam: PF00070; pyr\_redox; 1.  
DR PROSITE: PS00198; 4Fe4S\_FERREDOXIN; 1.  
KW Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S;  
KW Disease mutation.  
FT PROPEP 1 3  
FT CHAIN 4 1025 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+].  
FT NP\_BIND 335 351 NADP (POTENTIAL).  
FT NP\_BIND 471 481 FAD (POTENTIAL).  
FT DOMAIN 661 678 URACIL-BINDING (POTENTIAL).  
FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 963 963 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 966 966 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT VARIANT 29 29 C -> R (IN DYPD\*9A AND DYPD\*9B; LOSS OF  
FT ACTIVITY).  
FT /FTid-VAR\_005173. R -> W (IN DYPD\*8; LOSS OF ACTIVITY).  
FT /FTid-VAR\_005174. S -> N (IN DYPD\*4; LOW ACTIVITY).  
FT /FTid-VAR\_005175. I -> V (IN DYPD\*5).  
FT /FTid-VAR\_005176. R -> H (IN DYPD\*9B; 25% OF ACTIVITY).  
FT /FTid-VAR\_005177. V -> F (IN DYPD\*10; LOW ACTIVITY).  
FT /FTid-VAR\_005178. S -> N (IN REF. 2).  
FT CONFLICT 910 910 S -> N (IN REF. 2).  
FT SEQUENCE 1025 AA; 111374 MW; 521C9430C7F69AFA CRC64;  
QY 2 LPPPRKMKGLFSQAKISLF 20  
Db 176 LPPPEKMEAYS-AKIALF 193  
Query Match 52.0%; Score 52.5; DB 1; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.83;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
RESULT 12  
DPYD\_BOVIN STANDARD: PRT: 1025 AA.  
ID DPYD\_BOVIN  
AC 028007; GSTRV4;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Dihydropyrimidine dehydrogenase [NADP+] (EC 1.3.1.2) (DPD)  
DE (DHPDhase) (dihydrouracil dehydrogenase) (dihydrothymine  
DE dehydrogenase).  
GN DPYD.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RC MEDLINE=97069948; PubMed=8912928;  
RA Albin N., Johnson M.R., Diasio R.B.;  
RT "CDNA Cloning of bovine liver dihydropyrimidine dehydrogenase.";  
RL DNA Seq. 6:243-250(1996).  
RN [2]  
RP SEQUENCE OF 668-678, AND ENZYME REGULATION.  
RC TISSUE=Liver;  
RC MEDLINE=92041818; PubMed=1939061;  
RA Porter D.J.T., Chestnut W.G., Taylor L.C.E., Merrill B.M., Spector T.;  
RT "Inactivation of dihydropyrimidine dehydrogenase by 5-iodouracil.";  
RL J. Biol. Chem. 266:19988-19994(1991).  
CC -1- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE  
CC REDUCTION OF URACIL AND THYMINE.  
CC -1- CATALYTIC ACTIVITY: 5,6-dihydrouracil + NADP(+) -> uracil + NADPH.  
CC -1- COFACTOR: FAD AND FMN. ALSO CONTAINS TWO 4FE-4S CLUSTERS.  
CC -1- ENZYME REGULATION: INACTIVATED BY 5-iodouracil.  
CC -1- PATHWAY: INITIAL AND RATE-LIMITING STEP IN THE 3-STEP PATHWAY OF  
CC URACIL AND THYMIDINE CATABOLISM AND IN THE PATHWAY LEADING TO THE  
CC FORMATION OF BETA-ALANINE.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U20981; AAB40985.1; -;  
DR HSSP: P26886; 1ERD.  
DR InterPro: IPR001450; 4Fe4S\_ferredoxin.  
DR InterPro: IPR001295; DHO\_dh.  
DR InterPro: IPR001327; FAD\_pyr\_redox.  
DR InterPro: IPR003009; FMN\_enzyme.  
DR Pfam: PF01180; DHodehase; 1.  
DR Pfam: PF00037; fer4; 2.  
DR Pfam: PF00070; pyr\_redox; 1.  
DR PROSITE: PS00198; 4Fe4S\_FERREDOXIN; 1.  
KW Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S.  
FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).  
FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).  
FT SEQUENCE 1025 AA; 11696 MW; 5B55F93A06C47E4F CRC64;  
QY 2 LPPPRKMKGLFSQAKISLF 19  
Db 176 LPPPEKMEAYS-AKIAL 192  
Query Match 45.0%; Score 45.5; DB 1; Length 1025;  
Best Local Similarity 61.1%; Pred. No. 12;  
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
RESULT 13

```

RIR2_HSV11
ID RIR2_HSV11 STANDARD: PRT: 340 AA.
AC PI0224;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Ribonucleoside-diphosphate reductase small chain (EC 1.17.4.1)
DE (Ribonucleotide reductase) (38 kDa subunit).
GN UL40.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeech D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thioedoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thioedoxin.
CC -1- COFACTOR: BINDS 2 IRON IONS.
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: D10879; BAO1686.1; -
CC EMBL: X14112; CAA32303.1; -
CC DR PIR: D30088; WMBE87.
CC DR InterPro: IPR000358; RibonucL_reductse.
CC DR Pfam: PF00268; ribonuc_red_sm; 1.
CC DR PROSITE: PS00368; RIBORED_SMALL; 1.
CC KW Oxidoreductase; DNA replication; Iron.
CC FT METAL 94 94 IRON 1 (BY SIMILARITY).
CC FT METAL 124 124 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 127 127 IRON 1 (BY SIMILARITY).
CC FT METAL 187 187 IRON 2 (BY SIMILARITY).
CC FT METAL 221 221 IRON 2 (BY SIMILARITY).
CC FT METAL 224 224 IRON 2 (BY SIMILARITY).
CC FT ACT_SITE 131 131 BY SIMILARITY.
CC SEQUENCE 340 AA; 38019 MW; 4B4ED94BF74FD3F CRC64;

Query Match 43.6%; Score 44; DB 1; Length 340;
Best Local Similarity 61.5%; Pred. No. 6.6;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PPRKMKGLFSQA 15
DB 242 PPDPRYGLFRQA 254

```

```

DE (Ribonucleotide reductase) (38 kDa subunit).
GN UL40.
OS Herpes simplex virus (type 1 / strain KOS).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83059830; PubMed=6292456;
RA Draper K.G., Frink R.J., Wagner E.K.;
RT "Detailed characterization of an apparently unspliced beta herpes
RT simplex virus type 1 gene mapping in the interior of another."
RL J. Virol. 43:1123-1128(1982).
RN [2]
RP REVISIONS.
RA Wagner E.K.;
RX Submitted (MAY-1995) to the EMBL/GenBank/DDA databases.
CC -1- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thioedoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thioedoxin.
CC -1- COFACTOR: BINDS 2 IRON IONS.
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: J02212; AAA66436.1; -
CC DR InterPro: IPR000358; RibonucL_reductse.
CC DR Pfam: PF00268; ribonuc_red_sm; 1.
CC DR PROSITE: PS00368; RIBORED_SMALL; 1.
CC KW Oxidoreductase; DNA replication; Iron.
CC FT METAL 94 94 IRON 1 (BY SIMILARITY).
CC FT METAL 124 124 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 127 127 IRON 1 (BY SIMILARITY).
CC FT ACT_SITE 131 131 BY SIMILARITY.
CC SEQUENCE 340 AA; 37966 MW; 921DC04B9D278DE5 CRC64;

Query Match 43.6%; Score 44; DB 1; Length 340;
Best Local Similarity 61.5%; Pred. No. 6.6;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PPRKMKGLFSQA 15
DB 242 PPDPRYGLFRQA 254

RESULT 15
DNAL METWA
ID DNAL METWA STANDARD: PRT: 389 AA.
AC P35515;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chapterone protein dnal.
GN DNAL.
OS Methanosarcina mazel (Methanosarcina frisia).
OS Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S-6;
RX MEDLINE=93324351; PubMed=8332479;

```



RA Macario A.J.L., Dugan C.B., Clarens M., Conway de Macario E.:  
 RT "dnaJ in Archaea."  
 CC Nucleic Acids Res. 21:2773-2773(1993).  
 CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -1- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 CR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: X60265; CAA42813.1; -.  
 DR PIR: S41748; S41748.  
 DR HSSP: P25685; 1HDJ.  
 DR InterPro: IPR003095; DnaJ.  
 DR InterPro: IPR002939; DnaJ\_C.  
 DR InterPro: IPR001305; DnaJ\_CXXCXG.  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR Pfam: PF00226; DnaJ\_1.  
 DR Pfam: PF01556; DnaJ\_C\_1.  
 DR Pfam: PF00684; DnaJ\_CXXCXG\_1.  
 DR PRINTS: PR00625; DNADJPROTEIN.  
 DR SMART: SM00271; DnaJ\_1.  
 DR PROSITE: PS00636; DNADJ\_1; 1.  
 DR PROSITE: PS00076; DNADJ\_2; 1.  
 DR PROSITE: PS00637; DNADJ\_CXXCXG\_1.  
 KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
 FT DOMAIN 4 72  
 FT DOMAIN 110  
 FT REPEAT 144 151 GLY-RICH.  
 FT REPEAT 161 168 CXXCXG MOTIF.  
 FT REPEAT 187 194 CXXCXG MOTIF.  
 FT REPEAT 201 208 CXXCXG MOTIF.  
 FT METAL 144 144 ZINC 1 (BY SIMILARITY).  
 FT METAL 147 147 ZINC 1 (BY SIMILARITY).  
 FT METAL 161 161 ZINC 2 (BY SIMILARITY).  
 FT METAL 164 164 ZINC 2 (BY SIMILARITY).  
 FT METAL 187 187 ZINC 2 (BY SIMILARITY).  
 FT METAL 190 190 ZINC 2 (BY SIMILARITY).  
 FT METAL 201 201 ZINC 1 (BY SIMILARITY).  
 FT METAL 204 204 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 389 AA; 42990 MW; 5B840E379D8C8139 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 389;  
 Best Local Similarity 66.7%; Pred. No. 11;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 PRKMGLESOAK 16  
 DB 373 PRKSGLEFKVK 384

RESULT 16  
 ID MTS1\_STRAL STANDARD; PRT: 587 AA.  
 AC 053609:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Modification methylase Sali (EC 2.1.1.72) (Adenine-specific  
 DE methyltransferase Sali) (M.Sali).  
 GN SALIM.  
 OS Streptomyces albus C.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95129852; PubMed=7828868;  
 RA Rodicio M.R., Quinton-Jager T., Moran L.S., Slatko B.E., Wilson G.G.;  
 RT "Organization and sequence of the Sali restriction-modification  
 RT system."  
 RL Gene 151:167-172(1994).  
 CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
 CC GTGAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND  
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE SALI ENDONUCLEASE.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
 CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: U01232; AAA81887.1; -.  
 DR HSSP: P14385; 2ADM.  
 DR REBASE: 3491; M.Sali.  
 DR InterPro: IPR002296; N12N6\_mtfase.  
 DR InterPro: IPR002052; N6\_Mtase.  
 DR InterPro: IPR000051; SAM\_bind.  
 DR PRINTS: PR00507; N12N6MTFRASE.  
 DR PROSITE: PS00092; N6\_MTASE; 1.  
 KW transferase; Methyltransferase; Restriction system.  
 SQ SEQUENCE 587 AA; 64975 MW; B482F374397FF999 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 587;  
 Best Local Similarity 70.0%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VLPPPRKMG 10  
 DB 334 ILPSPRKMG 343

RESULT 17  
 ID HK26\_MOUSE STANDARD; PRT: 289 AA.  
 AC P43688:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Homeobox protein NKX-2.6.  
 DE NKX2F OR NKX2-6 OR NKX-2.6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALE/C;  
 RX MEDLINE=98213637; PubMed=9545560;  
 RA Biben C., Hatzistavrou T., Wang C.-C., Harvey R.P.;  
 RT "Expression of NK-2 class homeobox gene NKx2-6 in foregut endoderm  
 RT and heart."  
 RL Mech. Dev. 73:125-127(1998).  
 RN [2]  
 RP SEQUENCE OF 123-214 FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=94116443; PubMed=7904557;  
 RA Lints T.J., Parsons L.M., Hartley L., Lyons I., Harvey R.P.;  
 RT "Nkx-2.5: a novel murine homeobox gene expressed in early heart  
 RT progenitor cells and their myogenic descendants."  
 RL Development 119:419-431(1993).

```
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING GUT ENDODERM, CARDIAC
CC PROGENITORS AND HEART. ALSO EXPRESSED IN NEONATAL BRAIN.
CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF045150; AAC15674.1; -.
DR HSP: P23441; IFTT.
DR MGD: MGI:97351; NKx2-6.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox_1.
DR PRINTS: PR00024; HOMEBOX.
DR SMART: SM00389; HOX_1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR Homeobox: DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 123 182 HOMEBOX.
SQ SEQUENCE 289 AA; 31597 MW; DBF8740A39CC0805 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 289;
Best Local Similarity 57.1%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 PPRKMGILFSQAKI 17
Db 122 PPRKMGILFSQAKI 135

RESULT 18
MCAL_CRIGR STANDARD; PRT; 359 AA.
AC 054873;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multisynthetase complex auxiliary component p43 (Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE Inducible cytokine subfamily E member 1)).
GN SCYL.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98070438; PubMed=9405472;
RA Quesvillon S., Agou F., Robinson J.-C., Mirande M.;
RT "The p43 component of the mammalian multi-synthetase complex is likely
RT to be the precursor of the endothelial monocyte-activating polypeptide
RT II cytokine."
RL J. Biol. Chem. 272:32573-32579(1997).
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: AF021800; AB95207.1; -.
DR InterPro: IPR002547; tRNA_bind.
DR Pfam: PF01588; tRNA_bind_1.
KW Protein biosynthesis; tRNA-binding; Cytokine.
FT DOMAIN 204 297 tRNA BINDING.
SQ SEQUENCE 359 AA; 39601 MW; 4D868D1B5D72C23 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 359;
Best Local Similarity 53.8%; Pred. No. 22;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 PPRKMGILFSQAKI 17
Db 262 PPRKMGILFSQAKI 274

RESULT 19
AXOL_RAT STANDARD; PRT; 1040 AA.
AC P22063;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Contactin 2 precursor (Axonin-1) (Axonal glycoprotein TAG-1)
DE (Transferrin axonal glycoprotein 1) (TAX-1).
GN CNTN2 OR TAXI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC TISSUE=Spinal cord;
RX MEDLINE=90199890; PubMed=2317872;
RA Furely A.J., Motson S.B., Manalo D., Karagozeos D., Dodd J.,
RA Jessell T.M.;
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily
RT member with neurite outgrowth-promoting activity."
RL Cell 61:157-170(1990).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS, AND IN ADULT
CC BRAIN, SPINAL CORD AND CEREBELLUM.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED ON A SUBSET OF AXONS
CC IN THE DEVELOPING RAT NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31725; AAA42201.1; -.
DR PIR: A34695; A34695.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF00041; fn3; 4.
DR Pfam: PF00047; Ig; 6.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00408; IGG2; 5.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 30
```



```
CC SIMILARITY)
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS. STRONG, TO E.COLI ZNTR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32835; AAC23268.1; -
CC TIGR: H11623; -
CC InterPro: IPR000551; HTH_MERR.
CC DR Pfam: PF00376; merr; 1.
CC SMART: SM00422; HTH_MERR; 1.
CC DR PROSITE: PS00552; HTH_MERR_FAMILY; 1.
CC KM Transcription regulation: DNA-binding; Zinc: Complete proteome.
CC FT DNA BIND 4 23 H-T-H MOTIF (POTENTIAL).
CC SQ SEQUENCE 135 AA; 15636 MW; CE07D848D0BCF19C CRC64;

Query Match 39.6%; Score 40; DB 1; Length 135;
Best Local Similarity 42.9%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VLPPRRMKGLFSQ 14
   :|:::|:|:|
Db 25 LIPPRKRTSGNFRQ 38

RESULT 22
YOM1_CAEEL STANDARD; PRT; 292 AA.
ID YOM1_CAEEL
AC P30647;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DT Hypothetical 32.7 kDa protein ZK643.1 in chromosome III.
GN ZK643.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peleodermidae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=92168156; Pubmed=1538779;
RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,
RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,
RA Craxton M., Durbin R.K., Berts M., Metzstein M., Hawkins T.,
RA Ainscough R., Waterston R.;
RT "The C. elegans genome sequencing project: a beginning.";
RL Nature 356:37-41(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11126; CAA77470.1; -
CC PIR: S23239; S23239.
CC DR MornBep: ZK643.1; CE00440.
CC KM Hypothetical protein.
CC SQ SEQUENCE 292 AA; 32669 MW; 0EDCA0FB82ABFB6A CRC64;

Query Match 39.6%; Score 40; DB 1; Length 292;
Best Local Similarity 52.6%; Pred. No. 26;
```

```
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLPPRRMKGLFSQAKISL 19
   | | | | | | | |
Db 9 VLAPRCMGGEFFAKVILL 27

RESULT 23
MCAL_MOUSE STANDARD; PRT; 310 AA.
ID MCAL_MOUSE
AC P31230; Q60659;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE inducible cytokine subfamily E member 1)].
GN scyel OR EMAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014290; Pubmed=7929199;
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,
RA Gikschelt T., Chabot J., Nowygrod R., Greenberg S., Kang W.J.,
RA Leung D.W., Hayward J.R., Kistel W., Heath M., Brett J., Stern D.M.;
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II ";
RL J. Biol. Chem. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 145-164 FROM N.A.
RX MEDLINE=93015897; Pubmed=1400342;
RA Kao J., Ryan J., Brett G., Chen J., Shen H., Fan Y.-G., Godman G.,
RA Familletti P.C., Wang F., Pan Y.-C.E., Stern D., Claus M.;
RT "Endothelial monocyte-activating polypeptide II. A novel
RT tumor-derived polypeptide that activates host-response mechanisms ";
RL J. Biol. Chem. 267:20239-20247(1992).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94193665; Pubmed=7545917;
RA Kao J., Fan Y., Haehnel I., Brett J., Greenberg S., Claus M.,
RA Kayton M., Houck K., Kistel W., Seljelid R., Burnier J., Stern D.;
RT "A peptide derived from the amino terminus of endothelial-monocyte-
RT activating polypeptide II modulates mononuclear and polymorphonuclear
RT leukocyte functions, defines an apparently novel cellular interaction
RT site, and induces an acute inflammatory response.";
RL J. Biol. Chem. 269:9774-9782(1994).
CC -1- FUNCTION: ALTERS ENDOTHELIAL AND MONOCYTE FUNCTIONS. INDUCES THE
CC MIGRATION OF MONOCYTES AND GRANULOCYTES, AND INDUCES AN
CC INFLAMMATORY RESPONSE IN THE MOUSE FOOTPAD MODEL. EMAP II ELICITS
CC A PHLOGENIC RESPONSE AND, POTENTIALLY, AUGMENTS THE EFFECTS OF
CC THE OTHER TUMOR-DRIVEN CYTOKINES.
CC -1- SUBUNIT: MONOMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U10118; AAA62203.1; -
CC DR EMBL: BC002054; AAH02054.1; -
CC PIR: A44032; A44032.
CC DR MGD: MGI:102774; Emmap2.
CC DR InterPro: IPR002547; CRNA_bind.
```



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:52 ; Search time 204.58 Seconds  
(Without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-42  
Perfect score: 101  
Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP\_rvirus:\*  
16: SP\_bacteriophage:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	94	12 09ENW1	09enw1 hepatitis a
2	101	100.0	94	12 09ENW0	09enw0 hepatitis a
3	101	100.0	94	12 09ENW9	09enw9 hepatitis a
4	101	100.0	94	12 09ENW8	09enw8 hepatitis a
5	101	100.0	94	12 09ENW7	09enw7 hepatitis a
6	101	100.0	94	12 09ENW6	09enw6 hepatitis a
7	101	100.0	94	12 09ENW5	09enw5 hepatitis a
8	101	100.0	94	12 09ENW4	09enw4 hepatitis a
9	101	100.0	94	12 09ENW3	09enw3 hepatitis a
10	101	100.0	94	12 09ENW2	09enw2 hepatitis a
11	101	100.0	94	12 09ENW1	09enw1 hepatitis a
12	101	100.0	94	12 09ENW0	09enw0 hepatitis a
13	101	100.0	94	12 09ENW9	09enw9 hepatitis a
14	101	100.0	94	12 09ENW8	09enw8 hepatitis a
15	101	100.0	94	12 09ENW7	09enw7 hepatitis a
16	101	100.0	94	12 09ENW6	09enw6 hepatitis a

ALIGNMENTS

17	101	100.0	94	12 09ENW4	09enw4 hepatitis a
18	101	100.0	94	12 09ENW2	09enw2 hepatitis a
19	101	100.0	94	12 09ENW1	09enw1 hepatitis a
20	101	100.0	94	12 09ENW9	09enw9 hepatitis a
21	101	100.0	94	12 09ENW8	09enw8 hepatitis a
22	101	100.0	94	12 09ENW7	09enw7 hepatitis a
23	101	100.0	115	12 09DMR4	09dmr4 hepatitis a
24	101	100.0	116	12 092941	092941 hepatitis a
25	101	100.0	116	12 09W7S7	09w7s7 hepatitis a
26	101	100.0	116	12 09W7X7	09w7x7 hepatitis a
27	101	100.0	116	12 071977	071977 hepatitis a
28	101	100.0	126	12 098W19	098w19 hepatitis a
29	101	100.0	258	12 0991S7	0991s7 hepatitis a
30	101	100.0	1124	12 084780	084780 hepatitis a
31	101	100.0	1161	12 005794	005794 hepatitis a
32	101	100.0	2225	12 09DL32	09dl32 hepatitis a
33	101	100.0	2227	12 09WMA1	09wma1 hepatitis a
34	101	100.0	2227	12 09WM99	09wm99 hepatitis a
35	101	100.0	2227	12 091FH5	091fh5 hepatitis a
36	99	98.0	116	12 09W7X8	09w7x8 hepatitis a
37	98	97.0	94	12 09ENW0	09enw0 hepatitis a
38	98	97.0	94	12 09ENW3	09enw3 hepatitis a
39	98	97.0	94	12 09ENW8	09enw8 hepatitis a
40	98	97.0	94	12 09ENW7	09enw7 hepatitis a
41	98	97.0	94	12 09ENW6	09enw6 hepatitis a
42	98	97.0	94	12 09ENW4	09enw4 hepatitis a
43	98	97.0	94	12 09ENW3	09enw3 hepatitis a
44	98	97.0	112	12 086534	086534 hepatitis a
45	98	97.0	115	12 09DMR5	09dmr5 hepatitis a
46	98	97.0	116	12 09W7T4	09w7t4 hepatitis a
47	98	97.0	116	12 071975	071975 hepatitis a
48	98	97.0	125	12 0913V2	0913v2 hepatitis a
49	98	97.0	125	12 09WMA2	09wma2 hepatitis a
50	98	97.0	2216	12 067824	067824 hepatitis a
51	98	97.0	2218	12 067817	067817 hepatitis a
52	98	97.0	2218	12 067825	067825 hepatitis a
53	98	97.0	2227	12 067826	067826 hepatitis a
54	98	97.0	2227	12 09WMA3	09wma3 hepatitis a
55	98	97.0	94	12 09ENW2	09enw2 hepatitis a
56	97	96.0	94	12 09ENW0	09enw0 hepatitis a
57	97	96.0	94	12 09ENW9	09enw9 hepatitis a
58	97	96.0	94	12 09ENW8	09enw8 hepatitis a
59	97	96.0	94	12 09ENW7	09enw7 hepatitis a
60	97	96.0	2227	12 09WMA4	09wma4 hepatitis a

RESULT 1  
ID 09ENW1 PRELIMINARY: PRT: 94 AA.  
AC 09ENW1: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
OC [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=NAGASAKI 01;  
RA Iida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kikuchi Y.,  
Yasuno A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038276; BAB1814.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 94 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKLFQAKISLF 20  
DB 74 VLPPPRKMGKLFQAKISLF 93

## RESULT 2

Q9ENW0 PRELIMINARY; PRT; 94 AA.  
AC Q9ENW0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 02;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038277; BAB11815.1; -.  
FT NON\_TER 1 94  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10903 MW; 9AF8BE853D57329C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKLFQAKISLF 20  
DB 74 VLPPPRKMGKLFQAKISLF 93

## RESULT 3

Q9ENV9 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 03;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038278; BAB11816.1; -.  
FT NON\_TER 1 94  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10862 MW; 9AF9EFD4A8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;

Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKLFQAKISLF 20  
DB 74 VLPPPRKMGKLFQAKISLF 93

## RESULT 4

Q9ENV8 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 04;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038279; BAB11817.1; -.  
FT NON\_TER 1 94  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKLFQAKISLF 20  
DB 74 VLPPPRKMGKLFQAKISLF 93

## RESULT 5

Q9ENV7 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 06;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038280; BAB11818.1; -.  
FT NON\_TER 1 94  
FT NON\_TER 1 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

OY      1 VLPPPRKMKGLFSQAKISLF 20
        |||||
Db      74 VLPPPRKMKGLFSQAKISLF 93

```

RESULT		6		
09ENV6	09ENV6		PRELIMINARY;	PRT; 94 AA.
ID	09ENV6			
AC	09ENV6			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	POLYPROTEIN (FRAGMENT).			
OS	Hepatitis A virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;			
CC	Hepatovirus.			
OX	NCBI_TaxID=12092;			
BN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MGASAKI 07;			
RA	Ida S.;			
RT	*Influence of HIV-1 infection on acute hepatitis A virus (HAV)			
RT	infection: Prolonged HAV viremia and mild liver injury.*;			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB038281; BAB11819.1; ..			
FT	NON TER			
FT	1			
FT	94			
FT	94			
SO	SEQUENCE	94 AA;	10876 MW;	9AF8BE91BB8CABAC CXC64;

RESULT 7			
09ENV5	ID	09ENV5	PRELIMINARY; PRT: 94 AA.
AC	09ENV5:		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	POLYPROTEIN (FRAGMENT).		
OS	Hepatitis A virus.		
CC	viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.		
OX	NCBI_Taxid=12092;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NAGASAKI 08;		
RA	Ide S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y., Yasuoka A., Oka S.;		
RA	Influence of HIV-1 infection on acute hepatitis A virus (HAV) infection: Prolonged HAV viremia and mild liver injury.";		
RL	Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AB038282; BAB11820.1; -		
FT	NON_TER	1	
FT	NON_TER	1	
FT	NON_TER	94	
SO	SEQUENCE	94 AA; 10876 MW; 9AF8BBE91BB8C4BAC CRC64;	

	RESULT	8		
OGENV4				
ID	OGENV4	PRELIMINARY;	PRT:	94 AA.
AC	OGENV4			
DT	01-MAR-2001	(TREMBLrel. 16, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)		
DE	POLYPROTEIN (FRAGMENT).			
OS	Hepatitis A virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae.			
OX	Hepatovirus.			
RN	NCBI_TaxID=12092:			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RA	STRAIN-NAGASAKI 10;			
RT	Iida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kituchi Y.,			
RL	Yaenaka A., Oka S.;			
RT	"Influence of HIV-1 infection on acute hepatitis A virus (HAV)			
RT	infection: Prolonged HAV viremia and mild liver injury."			
DR	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.			
FT	EMBL; AB038283; BAB1821.1; -.			
FT	NON_TER			
SO	NON_TER			
SEQUENCE	94 AA.	10876 MW;	9AF8BE91BB8C4BAC	CRC64;

ID	OGENV3	PRELIMINARY;	PRT;	94 AA.
AC	OGENV3:			
DT	01-MAR-2001 (TREMBlrel. 16,	Created)		
DI	01-MAR-2001 (TREMBlrel. 16,	Last sequence update)		
DR	01-MAR-2001 (TREMBlrel. 16,	Last annotation update)		
DE	POLYPROTEIN (FRAGMENT).			
OS	Hepatitis A virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae.			
CC	Hepatovirus.			
OX	NCBI_TaxID=12092;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NAGASAKI 12N;			
RA	Iida S., Tsuchikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,			
RA	Yasuda A., Oka S.;			
RT	"Influence of HIV-1 infection on acute hepatitis A virus (HAV)			
RT	infection: Prolonged HAV viremia and mild liver injury.";			
RL	Submitted (Feb-2000) to the EMBL/Genbank/DDBJ databases.			
DR	EMBL: AB038284; BAB11822.f; -.			
FT	NON_TER	1	94	
FT	NON_TER	1	94	
SQ	SEQUENCE	94 AA;	10876 MW;	9AF8BBEBA1A366BAC CRC64;

	Query Match	100.0%	Score 101:	DB 12:	Length 94;
	Best Local Similarity	100.0%;	Pred. No. 1e-09;		
	Matches 20:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0.
QY	1 VLPPrKMKGLFSQAkISLF	20			
Db	74 VLPPrKMKGLFSQAkISLF	93			

Query Match	100.0%	Score 101;	DB 12;	Length 94;
Best Local Similarity	100.0%	Pred. No. 1e-09;		
Matches	20;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	1	VLP	PPR	KMKGLFSQAKISLF 20
DB	74	VLP	PPR	KMKGLFSQAKISLF 93
RESULT	10			
Q9JEN2				

ID Q9ENV2 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV2;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 15;  
RA Ida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038285; BAB11823.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGIFSQAISLF 20  
Db 74 VLPPPRKMGIFSQAISLF 93

RESULT 11  
Q9ENV1 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV1;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 18;  
RA Ida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038286; BAB11824.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGIFSQAISLF 20  
Db 74 VLPPPRKMGIFSQAISLF 93

RESULT 12  
Q9ENV9 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV9;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 21;  
RA Ida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038288; BAB11826.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGIFSQAISLF 20  
Db 74 VLPPPRKMGIFSQAISLF 93

RESULT 13  
Q9ENV8 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV8;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DE POLYPROTEIN (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAGASAKI 22;  
RA Ida S., Tachikawa N., Nakajima A., Dalkoku M., Yano M., Kikuchi Y.,  
RT Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB038289; BAB11827.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGIFSQAISLF 20  
Db 74 VLPPPRKMGIFSQAISLF 93

RESULT 14  
Q9ENV7 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV7;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).

```
OS Hepatitis A virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCB1_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NAGASAKI 23;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
  Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
  infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB038290; BAB11828.1; -.
FT NON_TER 1 94
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;
SQ

Query Match          100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPEPRKMKGLFSQAKISLF 20
DB 74 VLPPEPRKMKGLFSQAKISLF 93

RESULT 15
O9ENUT6 PRELIMINARY; PRT; 94 AA.
AC O9ENUT6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCB1_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NAGASAKI 24;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
  Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
  infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB038291; BAB11829.1; -.
FT NON_TER 1 94
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;
SQ

Query Match          100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPEPRKMKGLFSQAKISLF 20
DB 74 VLPPEPRKMKGLFSQAKISLF 93

RESULT 16
O9ENUT5 PRELIMINARY; PRT; 94 AA.
AC O9ENUT5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
```

```
OX NCB1_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NAGASAKI 25;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
  Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
  infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB038292; BAB11830.1; -.
FT NON_TER 1 94
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;
SQ

Query Match          100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPEPRKMKGLFSQAKISLF 20
DB 74 VLPPEPRKMKGLFSQAKISLF 93

RESULT 17
O9ENUT4 PRELIMINARY; PRT; 94 AA.
AC O9ENUT4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCB1_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NAGASAKI 26;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
  Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
  infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB038293; BAB11831.1; -.
FT NON_TER 1 94
FT SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;
SQ

Query Match          100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPEPRKMKGLFSQAKISLF 20
DB 74 VLPPEPRKMKGLFSQAKISLF 93

RESULT 18
O9ENUT2 PRELIMINARY; PRT; 94 AA.
AC O9ENUT2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCB1_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN-NAGASAKI 28;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL: AB038295; BAB11833.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;  
  
Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VLPPPRKMGLEFSQAKISLF 20  
DB 74 VLPPPRKMGLEFSQAKISLF 93  
|||||  
  
RESULT 19  
O9ENT5 PRELIMINARY; PRT; 94 AA.  
AC O9ENT1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 29;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL: AB038296; BAB11834.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AFBBE91BB8C4BAC CRC64;  
  
Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VLPPPRKMGLEFSQAKISLF 20  
DB 74 VLPPPRKMGLEFSQAKISLF 93  
|||||  
  
RESULT 20  
O9ENT9 PRELIMINARY; PRT; 94 AA.  
AC O9ENT1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NAGASAKI 32;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;

RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL: AB038298; BAB11836.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10807 MW; F194CE91BB8C4AFD CRC64;  
  
Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VLPPPRKMGLEFSQAKISLF 20  
DB 74 VLPPPRKMGLEFSQAKISLF 93  
|||||  
  
RESULT 21  
O9ENT5 PRELIMINARY; PRT; 94 AA.  
AC O9ENT5;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KANTOU 42;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
DR EMBL: AB038302; BAB11840.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 94 AA; 10842 MW; 741850915B86E1A6 CRC64;  
  
Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VLPPPRKMGLEFSQAKISLF 20  
DB 74 VLPPPRKMGLEFSQAKISLF 93  
|||||  
  
RESULT 22  
O9ENT1 PRELIMINARY; PRT; 94 AA.  
AC O9ENT1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KANTOU 49;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB038306; BAB1844.1; -  
FT NON\_TER 1  
RA 94  
SQ SEQUENCE 94 AA; 10842 MW; 741850915B86E1A6 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 94;  
Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMKGLFSQAKISLF 20  
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 23  
O9DMR4 PRELIMINARY; PRT; 115 AA.

AC O9DMR4; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-26;  
RA Steene-Johansen K.; Skaug K., Blystad H.;  
RX MEDLINE=20041342; PubMed=10574047;  
RT "overskating av hepatitt A ved molekylarepidemiologisk undersokelser";  
RL Tidskr. Nor laegeforen. 119:3725-3728(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-26;  
RA Steene-Johansen K.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ299461; CAC14071.2; -  
FT NON\_TER 1  
FT CHAIN 1 >64 VP1 PROTEIN.  
FT CHAIN 65 >115 P2A PROTEIN.  
FT NON\_TER 115  
SQ SEQUENCE 115 AA; 13344 MW; 16DFB8E7EB8A77B0 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 115;  
Pred. No. 1.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGLFSQAKISLF 115

RESULT 24  
O92941 PRELIMINARY; PRT; 116 AA.

AC O92941; 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-1;  
RX MEDLINE=97442186; PubMed=9298735;

RA Grinde B., Steene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RT Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75(1997).  
DR EMBL: AF050223; AAC05516.1; -  
DR EMBL: AF050231; AAC05524.1; -  
DR EMBL: AF050229; AAC05522.1; -  
DR EMBL: AF050230; AAC05523.1; -  
KW Polyprotein.  
FT NON\_TER 1  
FT NON\_TER 116  
SQ SEQUENCE 116 AA; 13479 MW; 4726C6A1E7EB8A64 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 116;  
Pred. No. 1.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGLFSQAKISLF 115

RESULT 25  
O9W7S7 PRELIMINARY; PRT; 116 AA.

AC O9W7S7; 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-4, AND NOR-3;  
RX MEDLINE=97442186; PubMed=9298735;  
RA Grinde B., Steene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RA Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-4, AND NOR-3;  
RA Steene-Johansen K., Skaug K., Blystad H., Grinde B.;  
RT "An unique hepatitis A virus strain caused an epidemic in Norway associated with intravenous drug abuse";  
RL Scand. J. Infect. Dis. 0:0-0(1998).  
DR EMBL: AF050226; AAC05519.1; -  
DR EMBL: AF050225; AAC05518.1; -  
KW Polyprotein.  
FT NON\_TER 1  
FT NON\_TER 116  
SQ SEQUENCE 116 AA; 13507 MW; 8E26DFB8E7EB8A77 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 116;  
Pred. No. 1.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGLFSQAKISLF 115

Search completed: June 16, 2002, 00:08:52  
Job time: 790 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:01:44 ; Search time 209.1 Seconds  
(Without alignments)  
10.624 Million cell updates/sec

Title: US-09-171-432a-43

Perfect score: 101  
Sequence: 1 FSOAKTSLFYPEHEIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	20	18	AAW42926	Immunogenic Hepati
2	101	100.0	21	22	AAW50231	Synthetic HAV P2A
3	101	100.0	854	6	AAW50287	Sequence encoded b
4	101	100.0	1077	20	AAW53559	A partial hepatitis
5	101	100.0	1091	14	AAW53559	translated from 5'
6	101	100.0	2227	11	AAW5697	Attenuated hepatit
7	101	100.0	2227	18	AAW54074	Hepatitis A virus
8	101	100.0	2227	21	AAW5607	Amino acid sequenc
9	101	100.0	2227	21	AAW5608	Amino acid sequenc
10	101	100.0	2227	21	AAW5609	Amino acid sequenc
11	98	97.0	2227	7	AAW60066	Sequence of viral

12	97	96.0	366	6	AAW50230	Sequence of hepati
13	97	96.0	993	6	AAW50116	Sequence of Hepati
14	97	96.0	993	6	AAW50231	Sequence encoded b
15	49.5	49.0	20	19	AAW76539	Plasmid pGEM2-HAV-
16	47	46.5	518	20	AAW27340	Group B Streptococ
17	47	46.5	816	20	AAW27339	Group B Streptococ
18	42	41.6	20	18	AAW42925	Immunogenic Hepati
19	42	41.6	21	22	AAW69442	Synthetic HAV P2A
20	42	41.6	503	22	ABG15762	Novel human diagno
21	41	40.6	144	22	AAW96238	Putative P. abyssi
22	41	40.6	173	22	AAW9279	Protein encoded by
23	41	40.6	238	22	ABW68889	Drosophila melanog
24	41	40.6	2431	13	AAW25138	SFV4 non-structura
25	40	39.6	105	21	AAW19719	Arabidopsis thalia
26	40	39.6	105	21	AAW61014	Arabidopsis thalia
27	40	39.6	122	21	AAW41997	Arabidopsis thalia
28	40	39.6	173	22	AAW29276	Human PRO polypept
29	40	39.6	173	22	AAW49278	Protein encoded by
30	40	39.6	681	22	ABW63032	Drosophila melanog
31	40	39.6	748	22	AAW58340	Human brain expres
32	39.5	39.1	126	22	AAW58545	Propionibacterium
33	39	38.6	148	20	AAW35073	Chlamydia pneumoni
34	39	38.6	170	20	AAW33564	G. max GBF homolog
35	39	38.6	200	22	ABW69162	Drosophila melanog
36	39	38.6	216	22	AAW98461	Callithrix jacchus
37	39	38.6	222	22	ABW08469	Novel human diagno
38	39	38.6	222	22	ABW15523	Human olfactory re
39	39	38.6	289	13	AAW30190	USTIE Exon 1 produ
40	39	38.6	289	21	AAW57095	UDP-glucuronosyltr
41	39	38.6	323	22	AAW71789	Human olfactory re
42	39	38.6	323	22	AAW72517	Human OR-1-like poly
43	39	38.6	397	22	AAW82075	S. epidermidis ope
44	39	38.6	455	20	AAW34954	Chlamydia pneumoni
45	39	38.6	476	22	ABW09560	Novel human diagno
46	39	38.6	536	22	ABW26969	Novel human diagno
47	39	38.6	571	22	ABW21709	Novel human diagno
48	39	38.6	670	22	AAW25612	Human protein sequ
49	39	38.6	729	22	AAW23730	Human EST encoded
50	39	38.6	750	15	AAW47858	Human LDL receptor
51	39	38.6	800	11	AAW07713	Human low density
52	39	38.6	839	22	AAW64837	Chronic hepatitis
53	39	38.6	839	22	AAW49601	Human low density
54	39	38.6	860	15	AAW47157	Sequence of human
55	39	38.6	860	15	AAW47860	Human LDL receptor
56	39	38.6	860	22	AAW90761	Human shear stress
57	39	38.6	872	22	ABW1799	Human LDL receptor
58	39	38.6	924	16	AAW78234	Chicken P95/human
59	39	38.6	1410	18	AAW07622	LDLR/Tf chimeric p
60	39	38.6	1418	22	AAW32831	Novel human secret

## ALIGNMENTS

RESULT 1	
AAW42926	
ID	AAW42926 standard; peptide: 20 AA.
XX	
AC	
XX	
DT	28-APR-1998 (first entry)
XX	
DE	Immunogenic Hepatitis A virus peptide YK-1319.
XX	
KW	Immunogenic peptide: Immunogenic epitope: P2A protein:
KM	Immune response; antibody.
XX	
OS	Synthetic.
XX	Hepatitis A virus.
XX	
FN	W09740147-A1.
XX	
PD	30-OCT-1997.





```

XX  Sequence      854 AA:
SQ
Query Match          100.0%; Score 101; DB 6; Length 854;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  1 FSOAKISLFTYEHEHIMKFS 20
    ||||||||||||||||
DB  834 fsgakislftyechehmkfs 853

RESULT 4
AAW95559
ID AAW95559 standard; Protein; 1077 AA.
AC AAW95559;
XX
XX 28-APR-1999 (first entry)
XX
XX A partial hepatitis A virus (HAV) protein.
XX
XX Hepatitis A virus protein; HAV; P2 region;
XX cell-culture-adapted HAV strain; infection; accelerated growth.
XX
XX Hepatitis A virus.
XX
XX US5849562-A.
XX
XX 15-DEC-1998.
XX
XX 06-JUN-1995; 95US-0468926.
XX
XX 06-NOV-1991; 91US-0788262.
XX 30-SEP-1983; 83US-0537911.
XX 27-SEP-1984; 84US-0554942.
XX 06-OCT-1988; 88US-0256135.
XX 06-JUN-1995; 95US-0468926.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Emerson SU, Purcell RH;
XX
XX WPI; 1999-094412/08.
XX N-PSDB; AAX01006.
XX
XX Chimeric hepatitis A virus strains - with P2 region from
XX cell-culture-adapted strain in wild-type genome
XX
XX Disclosure; Fig 7A-L; 36pp; English.
XX
XX The present sequence represents a partial hepatitis A virus (HAV)
XX protein. The specification describes a DNA construct consisting
XX of a wild-type HAV genome in which the P2 region is replaced by the
XX P2 region from a cell-culture-adapted HAV strain. The construct is
XX used to demonstrate that mutations in the P2 region of a
XX cell-culture-adapted HAV strain are sufficient for establishment of
XX infection and accelerated growth in cell culture.
XX
XX Sequence 1077 AA:
SQ
Query Match          100.0%; Score 101; DB 20; Length 1077;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  1 FSOAKISLFTYEHEHIMKFS 20
    ||||||||||||||||
DB  1057 fsgakislftyechehmkfs 1076

RESULT 5

```

```

AAR32426
ID AAR32426 standard; Protein; 1091 AA.
XX
XX AAR32426;
XX
XX 17-DEC-2001 (updated)
XX 10-JUN-1993 (first entry)
XX
XX Translated from 5' region of Hepatitis A virus genomic clone.
XX
XX HAV HM-175; chronic liver disease; picornavirus.
XX
XX Hepatitis A Virus.
XX
XX Key location/Qualifiers
XX Region 238..1091
XX /label= ORF
XX /note= "second putative initiation codon at
XX position 240"
XX Region 1..711
XX /note= "X's correspond to nonsense codons,
XX i.e. this region is not an ORF"
XX
XX USN788262-N.
XX
XX 15-DEC-1992.
XX
XX 30-SEP-1983; 83US-0536911.
XX
XX 27-SEP-1984; 84US-0654942.
XX 06-OCT-1988; 88US-0256135.
XX 30-SEP-1983; 83US-0536911.
XX 06-NOV-1991; 91US-0788262.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Baltimore D, Feinstone SM;
XX Purcell RH, Racanelli VR, Ticehurst JR;
XX
XX WPI; 1993-067429/08.
XX N-PSDB; AAO36934.
XX
XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.
XX of antigen and antibodies
XX
XX Disclosure; Fig 7; 65pp; English.
XX
XX HAV virion RNA was extracted from the livers of marmosets which had
XX been inoculated with HAV (the HAV had previously been passaged twice
XX in marmosets). The RNA was used to prepare ds cDNA clones by
XX standard methods. Clones contg. inserts which hybridised to RNA from
XX HAV-infected African Green Monkey kidney cells were selected for
XX further analysis. A 7.4kb restriction map (about 98% of the HAV
XX genome) was constructed from 5 overlapping inserts. The sequence of
XX the first 3.3kb (approx.) from the 5'-terminus was determined. An
XX amino acid sequence was deduced from the entire clone and an open
XX reading frame was identified starting at position 238. A comparison
XX of the predicted HAV amino acid sequences with the known capsid
XX protein sequences of other picornaviruses (poliovirus, foot and
XX mouth disease virus and encephalomyelitis virus) revealed areas of
XX local homology.
XX (Note: Revised entry submitted to correct the patent number format of
XX US Government-owned NRTS applications to prevent clashes with ongoing US
XX granted patent numbers. For further information please visit the Derwent
XX web site at www.derwent.com/dwpl/updates/nrtis-us.html.)
XX
XX Sequence 1091 AA:
SQ
Query Match          100.0%; Score 101; DB 14; Length 1091;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
OY      1 FSOAKISLFTEHEIMKFS 20
Db      1071 fsgakislftyeeheimkfs 1090

RESULT  6
AAR05697
ID      AAR05697 standard; protein; 2227 AA.
XX
AC      AAR05697;
XX
DE      15-AUG-1990 (first entry)
XX
DE      Attenuated hepatitis A virus.
XX
KW      Hepatitis A virus; vaccine; attenuated.
XX
OS      Hepatitis A virus, strain HM-175.
XX
FH      Key
FH      Location/Qualifiers
FT      Region
FT      1..23
FT      /label=VP4 = 1A
FT      Region
FT      24..245
FT      /label=VP2 = 1B
FT      Region
FT      246..491
FT      /label=VP3 = 1C
FT      Region
FT      492..791
FT      /label=VP1 = 1D
FT      Region
FT      792..980
FT      /label=2A
FT      Region
FT      981..1087
FT      /label=2B
FT      Region
FT      1088..1422
FT      /label=2C
FT      Region
FT      1423..1496
FT      /label=3A
FT      Region
FT      1497..1519
FT      /label=3B = VPg
FT      Region
FT      1520..1738
FT      /label=3C
FT      Region
FT      1739..2227
FT      /label=3D
XX
PN      US4894228-A.
XX
PD      16-JAN-1990.
XX
PF      12-JUL-1988; 88US-0217824.
XX
PR      12-JUL-1988; 88US-0217824.
PR      12-JUL-1988; 88US-0652967.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN.
XX
PI      Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;
PI      Daemer RJ, Gust ID;
XX
DR      WPI, 1990-075557/10.
DR      N-PSDB; AAO03512.
XX
PT      Vaccine against hepatitis A virus, infection - comprises novel
PT      attenuated hepatitis A virus strain.
XX
PS      Claim 1; Fig 1; 18pp; English.
XX
CC      The attenuated HAV is useful for inducing protective immunity against
CC      HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by
CC      several nucleotide changes distributed throughout the genome, is
CC      attenuated for chimpanzees, elicits serum neutralising antibodies, and is
CC      suitable for use as an HAV vaccine. It is noted that not all the changes
CC      are necessary for attenuation and use as a vaccine.
XX
SQ      Sequence 2227 AA;
```

```
Query Match      100.0%; Score 101; DB 11; Length 2227;
Best local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 FSOAKISLFTEHEIMKFS 20
Db      834 fsgakislftyeeheimkfs 853

RESULT  7
AAM34074
ID      AAM34074 standard; Protein; 2227 AA.
XX
AC      AAM34074;
XX
DE      27-APR-1998 (first entry)
XX
DE      Hepatitis A virus HM-175 protein sequence.
XX
KW      HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;
KW      infection; vaccine.
XX
OS      Hepatitis A virus HM-175.
XX
FH      Key
FH      Location/Qualifiers
FT      Protein
FT      1..23
FT      /label= VP4
FT      Protein
FT      24..245
FT      /label= VP2
FT      Protein
FT      246..491
FT      /label= VP3
FT      Protein
FT      492..791
FT      /label= VP1
FT      Protein
FT      792..980
FT      /label= 2A
FT      Protein
FT      981..1087
FT      /label= 2B
FT      Protein
FT      1088..1422
FT      /label= 2C
FT      Protein
FT      1423..1496
FT      /label= 3A
FT      Protein
FT      1497..1519
FT      /label= 3B
FT      Protein
FT      1520..1738
FT      /label= 3C
FT      Protein
FT      1739..2227
FT      /label= 3D
XX
PN      WO9740166-A2.
XX
PD      30-OCT-1997.
XX
PE      18-APR-1997; 97WO-US06506.
XX
PR      19-APR-1996; 96US-0015642.
XX
PA      (USSH ) US SEC DEPT HEALTH.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Emerson SU, Purcell RH, Raychaudhuri G;
PI      Emerson SU, Purcell RH, Raychaudhuri G;
XX
DR      WPI, 1997-535850/49.
DR      N-PSDB; AAT93023.
XX
XX      Human attenuated HAV genome containing simian HAV 2C gene - useful
PT      as vaccines against HAV infection
XX
PS      Disclosure; Fig 13A-D; 66pp; English.
XX
CC      This protein sequence is encoded by the human hepatitis A virus
CC      (HAV) HM-175 wild-type genome (see AAT93023). Attenuated strain
```

CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
CC kidney cells. A claimed DNA construct (1) comprises a genome of  
CC HAV, where the genome is a human attenuated HAV genome in which a  
CC region of the 2C gene has been replaced by a corresponding region  
CC from a 2C gene of a simian AGM-27 HAV genome (see AAT93024). The  
CC region of the 2C gene from AGM-27 contained in the construct  
CC preferably encodes amino acids 120-328 of the 2C protein, amino  
CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
CC transcript of (1); (2) a cell transfected with (1) or the RNA  
CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
CC its RNA transcript, can be used as a vaccine for preventing HAV in  
CC a mammal. (1) or the RNA transcript can also be used to stimulate  
CC the production of protective antibodies in the mammal.

SO Sequence 2227 AA:

Query Match 100.0%; Score 101; DB 18; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYEEHEIMKFS 20  
|||||  
DB 834 fsgakislfyeeheimkfs 853

## RESULT 8

ID AAB18607 standard; Protein; 2227 AA.  
XX

AC AAB18607;

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.

OS Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PF 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93WO-US08610.

PR 10-MAR-1995; 95US-0397232.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI; 2000-586464/55.

DR N-PSDB; AAA75476.

PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

PS line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX Disclosure; Fig 6A-K; 72pp; English.

CC The present sequence is derived from a wild type hepatitis A virus

CC (HAV) strain HM-174. The sequence is modified to produce HAV which

CC are adapted to growth in the human fibroblast-like cell line MRC-5.

CC The HAV is able to propagate in MRC-5 cells and retain appropriate

CC attenuation. It is useful as a live vaccine for prophylaxis of

CC hepatitis A in humans and other primates.

CC

XX

SO Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYEEHEIMKFS 20  
|||||  
DB 834 fsgakislfyeeheimkfs 853

## RESULT 9

ID AAB18608 standard; Protein; 2227 AA.  
XX

AC AAB18608;

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

XX P-35 virus.

OS Hepatitis A virus.

PN US6113912-A.

PD 05-SEP-2000.

PF 07-JUN-1995; 95US-0475886.

PR 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93WO-US08610.

PR 10-MAR-1995; 95US-0397232.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

DR WPI; 2000-586464/55.

DR N-PSDB; AAA75477.

PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

PS line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX Disclosure; Columns 67-78; 72pp; English.

CC The present sequence is derived from passage 35 of a wild type

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is

CC designated P-35 virus. The sequence is modified to produce HAV which

CC are adapted to growth in the human fibroblast-like cell line MRC-5.

CC The HAV is able to propagate in MRC-5 cells and retain appropriate

CC attenuation. It is useful as a live vaccine for prophylaxis of

CC hepatitis A in humans and other primates.

CC

XX

SO Sequence 2227 AA;

```

XX 15-JAN-2001 (first entry)
DT
XX Amino acid sequence of live attenuated Hepatitis A virus 4380.
DE
XX HAV, strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KM HAV 4380.
XX
XX Hepatitis A virus.
OS
XX US6113912-A.
PN
XX
XX 05-SEP-2000.
PD
XX
XX 07-JUN-1995; 95US-0475886.
PF
XX 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93MO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX MPI: 2000-586464/55.
DR N-PSDB; AAA75478.
XX
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
XX Disclosure: Columns 93-104; 72pp; English.
PS
XX
XX The present sequence is derived from a live attenuated hepatitis A
CC virus (HAV) of the invention, designated HAV 4380. The sequence is
CC produced by modifying wild type HAV strain HM-174. The HAV of the
CC invention are adapted to growth in the human fibroblast-like cell
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis
CC of hepatitis A in humans and other primates.
XX
XX Sequence 2227 AA;
SQ
Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FSOAKISLFYTEHEHMKFS 20
DB 834 fsgakislftyeehehmks 853

```

```

RESULT 11
AAP60066
ID AAP60066 standard; Protein; 2227 AA.
XX
XX AAP60066;
AC
XX
XX 26-JUN-1991 (first entry)
DT
XX
XX Sequence of viral L434 polypeptide encoded by the complete
DE nucleotide sequence of the HAV genome.
XX
XX Diagnosis; vaccine; passive immunotherapy.
KM
XX
XX Hepatitis A virus.
OS
XX
XX Key Location/Qualifiers
FH 1..245
FT Region /label= P1.1A
FT 246..491
FT Region /label= 1B

```

```

FT Region 492..836
FT /label= 1C
FT Region 837..980
FT /label= P2.2A
FT 981..1076
FT /label= 2B
FT 1077..1422
FT /label= 2C
FT 1423..1484
FT /label= P3.3A
FT 1485..1507
FT /label= 3B
FT 1508..1678
FT /label= 3C
FT 1679..2227
FT /label= 3D
XX
XX EP199480-A.
PN
XX
XX 29-OCT-1986.
PD
XX
XX 03-APR-1986; 86EP-0302465.
PF
XX
XX 03-APR-1985; 85US-0719329.
PR
XX
XX (CHIR-) CHIRON CORP.
PA
XX
XX Dina D, Potter SJ, Vannest GA, Caput D;
XX MPI: 1986-286213/44.
DR N-PSDB; AAN60080.
XX
XX Hepatitis A virus nucleotide sequence and polypeptide - and use
PT in prodn. of vaccines and diagnostic probes
PT
XX
XX Claim 5; Fig 1; 18pp; English.
PS
XX
XX AAN60080 and oligonucleotide fragments are useful in detection of
CC hepatitis A virus; transformed hosts may be used for expression of
CC polypeptides and fragments useful in vaccines without risk of
CC infection by the virus or in prodn. of particles which are capable
CC of inducing immunocompetent B cells for passive immunotherapy. Pref.
CC epitope is derived from AAs 445-657 or 792-848 of the HAV
CC polypeptide sequence (AAP60066).
XX
XX Sequence 2227 AA;
SQ
Query Match 97.0%; Score 98; DB 7; Length 2227;
Best Local Similarity 95.0%; Pred. No. 3.3e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FSOAKISLFYTEHEHMKFS 20
DB 834 fsgakislftydehehmks 853

```

```

RESULT 12
AAP50230
ID AAP50230 standard; Protein; 366 AA.
XX
XX AAP50230;
AC
XX
XX 28-NOV-1991 (first entry)
DT
XX
XX Sequence of hepatitis A virus (HAV) surface protein (VP-1).
DE
XX
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;
KM diagnostic assay.
XX
XX Hepatitis A virus.
OS
XX
XX EP138704-A.
PN

```

```
XX 24-APR-1985.
PD
XX
XX 09-OCT-1984; 84EP-0402025.
XX
XX 02-MAR-1984; 84US-0585942.
PR 14-OCT-1983; 83US-0541836.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Hughes JV, Scolnick EM, Tomassini JE;
XX
XX WPI: 1985-100818/17.
DR N-PSDB: AAN50274.
XX
XX New hepatitis A virus surface protein - useful for binding to
PT neutralising antibodies to the virus
XX
XX Claim 21; Page 46-48; 49pp; English.
PS
XX VPI is isolated by solubilisation of the intact virus in an aq.
CC anionic surfactant and a reducing agent. The viral proteins are sepd.
CC and the protein of molecular wt. 33000 daltons is sepd.
CC
XX
SQ Sequence 366 AA:

Query Match 96.0%; Score 97; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 6,4e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKF 19
   ||||||||||||||||
DB 343 fsqakislftyeehlmkf 361

RESULT 13
AAP50116
ID AAP50116 standard; Protein; 993 AA.
XX
XX AAP50116;
AC
XX 30-SEP-1991 (first entry)
DT
XX
XX Sequence of Hepatitis A virus (HAV) immunogenic peptides
DE VP-1, VP-2, VP-3 and VP-4.
DE
XX Antigenic protein; immunogen; vaccine.
KW
XX Hepatitis A virus (strain CR326).
OS
XX
XX EPI54587-A.
PN
XX
XX 11-SEP-1985.
PD
XX
XX 27-FEB-1985; 85EP-0400369.
PF
XX
XX 02-MAR-1984; 84US-0585818.
PR
XX
XX (MERI ) MERCK & CO INC.
PA
XX
XX Linemeyer DL, Menke JG, Reuben RG, Mitra SW;
XX
XX WPI: 1985-224964/37.
DR N-PSDB: AAN50139.
XX
XX New nucleotide sequences coding for hepatitis A virus antigens -
PT useful for eliciting normal immune response and in vaccines for
PT protecting against the virus
XX
XX Example; Page 11-17; 32pp; English.
PS
XX Within the sequence in AAN50139 is encoded the information necessary
CC
```

```
CC to make the antigenic proteins of HAV. The sequences encoding for
CC the structural proteins began at base 403. The key sub-unit
CC sequences within VP-1, designated Sequences I,II,III,IV, and V,
CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other
CC nucleotide sequences which are valuable as encoding antigenic
CC proteins are the sequences from base 1749 to base 2722; from base
CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from
CC base 1749 to base 2722 is esp. valuable as a vector for producing
CC antigen protein. Sequences II-V are claimed. X in AAP50116 denotes the
CC translation of a stop codon.
CC
XX
SQ Sequence 993 AA:

Query Match 96.0%; Score 97; DB 6; Length 993;
Best Local Similarity 100.0%; Pred. No. 1,9e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKF 19
   ||||||||||||||||
DB 970 fsqakislftyeehlmkf 988

RESULT 14
AAP50231
ID AAP50231 standard; Protein; 993 AA.
XX
XX AAP50231;
AC
XX 28-NOV-1991 (first entry)
DT
XX
XX Sequence encoded by partial sequence of hepatitis A virus (HAV),
DE including surface protein (VP-1).
DE
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;
KW diagnostic assay.
KW
XX Hepatitis A virus.
OS
XX
XX Key Location/Qualifiers
FH 628,993
FT Protein /note="claimed; X denotes translated stop codons
FT and unspecified triplets"
FT
XX
XX EPI38704-A.
PN
XX
XX 24-APR-1985.
PD
XX
XX 09-OCT-1984; 84EP-0402025.
PF
XX
XX 02-MAR-1984; 84US-0585942.
PR 14-OCT-1983; 83US-0541836.
XX
XX (MERI ) MERCK & CO INC.
PA
XX
XX Hughes JV, Scolnick EM, Tomassini JE;
XX
XX WPI: 1985-100818/17.
DR N-PSDB: AAN50274.
XX
XX New hepatitis A virus surface protein - useful for binding to
PT neutralising antibodies to the virus
XX
XX Disclosure; Page 17-23; 49pp; English.
PS
XX VPI is isolated by solubilisation of the intact virus in an aq.
CC anionic surfactant and a reducing agent. The viral proteins are sepd.
CC and the protein of molecular wt. 33000 daltons is sepd.
CC
XX
SQ Sequence 993 AA:

Query Match 96.0%; Score 97; DB 6; Length 993;
```



```

XX PN WO9942588-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-CA00114.
XX PR 20-FEB-1998; 98US-0075425.
XX PA (BIOC-) BIOCHEM VACCINS INC.
XX PI Boyer M, Brodeur BR, Charlebois I, Hamel J, Martin D;
XX PI Rioux C;
XX DR WPI; 1999-540309/45.
XX DR N-PSDB; AAX91103.
XX PT Novel group B Streptococcus antigens - useful as vaccine
XX PT compositions for prophylaxis or therapy of Streptococcus infections
XX PS Claim 26; Fig 1E; 154pp; English.
XX CC The invention provides Group B Streptococcus (GBS) antigens
XX CC (AAY27336-370) and nucleic acids (AAX91103-X91111) encoding the antigens.
XX CC The GBS antigens can be recombinantly expressed using standard
XX CC recombinant methodology. The GBS antigens of the invention can be used as
XX CC vaccine components for the treatment or prophylaxis of diseases and
XX CC symptoms mediated by Streptococcus infection, especially group A
XX CC Streptococcus (S. pyogenes), GBS or S. agalactiae, S. dysgalactiae, S.
XX CC uberis, S. norcristi, as well as Staphylococcus aureus. The vaccines are
XX CC administered to those individuals at risk of GBS infection, particularly
XX CC pregnant women and infants for sepsis, meningitis, and pneumonia, as well
XX CC as immunocompromised individuals, such as those with diabetes, liver
XX CC disease or cancer. The vaccines also have veterinary applications, such as
XX CC for the treatment of mastitis in cattle. The present sequence represents
XX CC a GBS antigen of the invention.
XX SQ Sequence 816 AA:

Query Match 46.5%; Score 47; DB 20; Length 816;
Best Local Similarity 45.0%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FSQAKISLFYTEHEIMKFS 20
DB 669 fggapialvknhehevata 688

RESULT 18
AAM42925
ID AAM42925 standard; peptide; 20 AA.
XX AC AAM42925;
XX DT 28-APR-1998 (first entry)
XX DE Immunogenic Hepatitis A virus peptide YK-1318.
XX KM Immunogenic Hepatitis A virus peptide YK-1318.
XX KM Immune response; antibody.
XX OS Synthetic.
XX OS Hepatitis A virus.
XX PN WO9740147-A1.
XX PD 30-OCT-1997.
XX PF 18-APR-1997; 97WO-US06891.
XX PR 19-APR-1996; 96US-0015644.

```

```

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Fields HA, Khudyakov YE;
XX DR WPI; 1997-535831/49.
XX PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an
XX PT immune response to HAV in a mammal or to detect the presence of
XX PT antibodies against HAV in a mammal
XX PS Claim 18; Page 112; 140pp; English.
XX CC Peptides AAM42922-30 are immunogenic peptides corresponding to
XX CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are
XX CC substantially similar to a portion of the amino acid sequence of the P2A
XX CC protein of HAV corresponding to amino acids 792-980. The present peptide
XX CC is derived from amino acids 823-842, and has a reactivity of 31.3% with
XX CC acute sera. Compositions containing the peptides can be used to induce an
XX CC immune response to HAV in a mammal. The peptides can also be used to
XX CC detect the presence of antibodies against HAV in mammalian serum. The
XX CC peptides can also be used to make an antibody against HAV by
XX CC administering the peptide to a mammal.
XX SQ Sequence 20 AA:

Query Match 41.6%; Score 42; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSQAKISLF 9
DB 12 fsqakislif 20

RESULT 19
AAB69442
ID AAB69442 standard; Peptide; 21 AA.
XX AC AAB69442;
XX DT 20-APR-2001 (first entry)
XX DE Synthetic HAV P2A peptide, SEQ ID NO: 42.
XX KM Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;
XX KM antigen; major structural capsid polypeptide; HAV antibody detection.
XX OS Synthetic.
XX OS Hepatitis A virus.
XX PN WO200105824-A2.
XX PD 25-JAN-2001.
XX PF 14-JUL-2000; 2000WO-US19267.
XX PR 15-JUL-1999; 99US-0144412.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Fields HA, Khudyakov YE;
XX DR WPI; 2001-112681/12.
XX PT Synthetic peptides used as antigen sources for enzyme immunoassays
XX PT detecting anti-hepatitis A virus and as vaccines -
XX PS Claim 13; Page 95; 130pp; English.
XX CC The present sequence is one of a number of synthetic peptides which are
XX CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides
XX CC comprise antigenic epitopes of the major structural capsid polypeptides

```

CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.

SQ Sequence 21 AA;

Query Match 41.6%; Score 42; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLF 9  
| | | | |  
Db 12 fsqakisl 20

RESULT 20

ABG15762 ID ABG15762 standard; Protein: 503 AA.

AC ABG15762;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15753.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

FA (HXSE-) HXSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS79949.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 46121; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 503 AA;

Query Match 41.6%; Score 42; DB 22; Length 503;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 SQAKISLFYTEHEH 16  
| | | | |  
Db 173 sdgkrlfytysel 187

RESULT 21

AAB96238 ID AAB96238 standard; Protein: 144 AA.

AC AAB96238;

DT 29-OCT-2001 (first entry)

DE Putative P. abyssi glycine decarboxylase complex H subunit.

KW Hyperthermophilic archaeon; hyperthermophilic protein.

OS Pyrococcus abyssi.

PN FR2792651-A1.

PD 27-OCT-2000.

PF 21-APR-1999; 99FR-0005034.

PR 21-APR-1999; 99FR-0005034.

PA (CNRS ) CNRS CENT NAT RECH SCI.

PA (IFRE-) IFREMER INST FR RECH EXPL MER.

PI Forterre P, Thiery JC, Prieur D, Dietrich J, Lecompte O;

PI Querellou J, Weissenbach J, Saurin W, Hellig R;

DR WPI: 2001-126236/14.

PT New nucleotide sequences isolated from Pyrococcus abyssi encode

PS proteins useful in industry -

CC The present invention relates to the genomic sequence of Pyrococcus  
CC abyssi (see AAF66431 and AHA41223-7) and P. abyssi proteins. P. abyssi is  
CC a hyperthermophilic archaeon, which is isolated from deep-sea  
CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
CC The proteins of the present invention have various potential industrial  
CC uses, since the proteins are stable at very high temperatures, some up to  
CC 110 degrees centigrade.

CC Note: This patent is in the same patent family as WO200065062, which  
CC contains additional sequences as shown in AAB99132-AAB99143,  
CC AAH75903-AAH75920 and AAG66436.

SQ Sequence 144 AA;



Query Match 40.6%; Score 41; DB 22; Length 144;  
Best Local Similarity 53.8%; Pred. No. 39;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 QAKISLFYTEHEHE 15  
DB 18 kvkeglyltyckene 30

RESULT 22

AA849279  
ID AAB49279 standard; Protein: 173 AA.

AA849279;

19-MAR-2001 (first entry)

Protein encoded by mouse ortholog of zsig81 DNA.

Cytokine; zsig 81; wound healing; proliferation; differentiation;

migration; metabolism.

Mus musculus.

WO200073459-A1.

07-DEC-2000.

01-JUN-2000; 2000WO-US15002.

01-JUN-1999; 99US-0323582.

(ZYMO) ZYMOGENETICS INC.

Piddington CS, West JR, Holly RD, Burkhead SK;

WPI; 2001-061540/07.

New zsig81 polypeptides and polynucleotides useful for e.g. promoting wound healing, or in diagnosing or treating disorders associated with cell loss or abnormal cell proliferation, such as cancer -

Disclosure: Page 97; 1099p; English.

The present invention relates to zsig81 and fragments thereof. The invention is useful for promoting wound healing, for modulating the cell proliferation, differentiation, migration or metabolism of responsive cell types that includes both primary and cultured cell lines, and for stimulating the proliferation of cells expressing markers associated with dendritic lineage cells.

Sequence 173 AA:

Query Match 40.6%; Score 41; DB 22; Length 173;  
Best Local Similarity 40.0%; Pred. No. 48;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEHEIMKFS 20

DB 148 waqslslfqlrhdimsr 167

RESULT 23

AB868889  
ID ABB68889 standard; Protein: 238 AA.

AB868889;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 33459.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL12992.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure: SEQ ID NO 33459; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL161840-ABL16175) and the encoded proteins

(ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_Pct\_sequences.

Sequence 238 AA:

Query Match 40.6%; Score 41; DB 22; Length 238;  
Best Local Similarity 44.4%; Pred. No. 68;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 SQAKISLFYTEHEHEIMKF 19

DB 201 smlslekylteehpckkf 218

RESULT 24

AA825138  
ID AAR25138 standard; Protein: 2431 AA.

AA825138;

05-JAN-1993 (first entry)

SFV4 non-structural polypeptide.

Semliki forest virus; SFV; SP6 expression vector; RNA polymerase.

Synthetic.

WO9210578-A.

25-JUN-1992.

12-DEC-1991; 91WO-SE00855.

13-DEC-1990; 90SE-0003978.

XX (BIOP-) BIOPTION AB.  
XX Garoff H, Liljestrom P;  
XX WPI, 1992-234633/28.  
DR N-PSDB: AAQ26021.  
XX  
XX RNA mol. derived from alphavirus RNA genome - chimeric alphavirus  
PT antigen and vaccine for immunisation against viral infections  
XX  
XX Disclosure: Fig 5; 94pp: English.  
XX  
XX The sequence given is the non-structural polypeptide encoded by a full  
CC length Semliki forest virus (SFV) cDNA clone within an SP6 expression  
CC vector. The SP6 RNA polymerase promoter allows in vitro transcription  
CC of full length and infectious transcripts.  
XX  
XX Sequence 2431 AA;  
SO

Query Match 40.6%; Score 41; DB 13; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 9e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 FSQAKISLFYREH 14  
| | : | : | : |  
Db 1072 fsapkslyennh 1085

RESULT 25  
AAC19719  
ID AAC19719 standard; Protein: 105 AA.  
XX AAC19719;  
AC  
XX 17-Oct-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21625.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140821.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.	Query Match 39.6%; Score 40; DB 21; Length 105;		
PR	05-AUG-1999;	99US-0147192.	Best Local Similarity 35.0%; Pred. No. 40;		
PR	05-AUG-1999;	99US-0147260.	Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;		
PR	06-AUG-1999;	99US-0147303.	Oy	1	FSQAKISLFTTEHRIKMS 20
PR	06-AUG-1999;	99US-0147416.	Db	26	fspaelstfslehnhfrln 45
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
PR	21-OCT-1999;	99US-0160741.			
PR	21-OCT-1999;	99US-0160767.			
PR	21-OCT-1999;	99US-0160768.			
PR	21-OCT-1999;	99US-0160770.			
PR	21-OCT-1999;	99US-0160814.			
PR	21-OCT-1999;	99US-0160815.			

Search completed: June 16, 2002, 00:01:45  
Job time: 12734 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 16, 2002, 00:03:16 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-43

Perfect score: 101

Sequence: 1 FSOAKISLFYTEHEIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

1: Issued\_Patents\_AA.\*  
2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	1091	6 5516630-2	Patent No. 5516630
2	101	100.0	2227	3 US-08-475-886-2	Sequence 2, Appl1
3	101	100.0	2227	3 US-08-475-886-4	Sequence 4, Appl1
4	101	100.0	2227	3 US-08-475-886-6	Sequence 6, Appl1
5	101	100.0	2227	4 US-08-397-232-2	Sequence 2, Appl1
6	101	100.0	2227	4 US-08-397-232-4	Sequence 4, Appl1
7	101	100.0	2227	4 US-09-171-387-2	Sequence 2, Appl1
8	41	40.6	2431	1 US-07-920-281C-2	Sequence 2, Appl1
9	41	40.6	2431	4 US-08-466-277-2	Sequence 2, Appl1
10	39	38.6	289	5 PCT-US92-00282-11	Sequence 11, Appl1
11	39	38.6	860	1 US-08-092-817-4	Sequence 4, Appl1
12	39	38.6	1410	2 US-08-470-058-4	Sequence 4, Appl1
13	39	38.6	1410	3 US-09-037-188-4	Sequence 4, Appl1
14	39	38.6	1410	4 US-09-285-310-4	Sequence 4, Appl1
15	38	37.6	265	1 US-07-958-551-2	Sequence 2, Appl1
16	38	37.6	265	1 US-08-129-610-7	Sequence 7, Appl1
17	38	37.6	265	1 US-08-129-609A-7	Sequence 7, Appl1
18	38	37.6	265	1 US-08-455-313-7	Sequence 7, Appl1
19	38	37.6	265	1 US-08-475-924-2	Sequence 2, Appl1
20	38	37.6	265	2 US-08-657-579A-2	Sequence 2, Appl1
21	38	37.6	265	4 US-09-224-025-7	Sequence 7, Appl1
22	38	37.6	265	5 PCT-US94-07887-7	Sequence 7, Appl1
23	38	37.6	707	3 US-08-772-270A-4	Sequence 2, Appl1
24	37	36.6	198	1 US-08-097-827-2	Sequence 2, Appl1
25	37	36.6	198	1 US-08-494-574-2	Sequence 2, Appl1
26	37	36.6	420	3 US-08-685-871-58	Sequence 58, Appl1
27	37	36.6	711	3 US-08-772-270A-12	Sequence 12, Appl1

28	37	36.6	1354	3 US-08-685-871-2	Sequence 2, Appl1
29	37	36.6	1388	2 US-08-685-576-1	Sequence 1, Appl1
30	37	36.6	1388	2 US-08-685-576-4	Sequence 4, Appl1
31	36	35.6	147	3 US-08-946-329A-62	Sequence 62, Appl1
32	36	35.6	169	2 US-08-912-227-2	Sequence 2, Appl1
33	36	35.6	416	3 US-08-946-329A-17	Sequence 17, Appl1
34	36	35.6	416	4 US-08-567-357A-17	Sequence 17, Appl1
35	36	35.6	416	4 US-08-729-733A-17	Sequence 17, Appl1
36	36	35.6	416	4 US-08-349-498-17	Sequence 17, Appl1
37	36	35.6	416	5 PCT-US95-15663-17	Sequence 17, Appl1
38	36	35.6	416	5 PCT-US95-15623-17	Sequence 17, Appl1
39	35.5	35.1	971	1 US-08-446-038B-19	Sequence 19, Appl1
40	35.5	35.1	971	1 US-08-446-010B-19	Sequence 19, Appl1
41	35.5	35.1	971	2 US-08-805-445-19	Sequence 19, Appl1
42	35.5	35.1	971	2 US-08-064-067D-19	Sequence 19, Appl1
43	35.5	35.1	971	2 US-09-066-208-19	Sequence 19, Appl1
44	35.5	35.1	993	1 US-08-446-010B-25	Sequence 25, Appl1
45	35.5	35.1	993	4 US-08-980-080-4	Sequence 4, Appl1
46	35.5	35.1	1129	1 US-08-357-598-6	Sequence 6, Appl1
47	35.5	35.1	1129	1 US-08-097-997A-9	Sequence 9, Appl1
48	35.5	35.1	1129	2 US-08-567-508C-3	Sequence 3, Appl1
49	35.5	35.1	1129	2 US-09-003-289-6	Sequence 6, Appl1
50	35.5	35.1	1129	3 US-09-196-480-3	Sequence 3, Appl1
51	35.5	35.1	1129	4 US-08-665-574C-9	Sequence 9, Appl1
52	35.5	35.1	1129	4 US-08-946-994-9	Sequence 9, Appl1
53	35.5	35.1	1129	5 PCT-US95-16435-6	Sequence 6, Appl1
54	35	34.7	135	3 US-08-050-259B-16	Sequence 16, Appl1
55	35	34.7	625	2 US-08-532-547-7	Sequence 7, Appl1
56	35	34.7	625	2 US-08-532-547-9	Sequence 9, Appl1
57	35	34.7	625	4 US-09-019-809-7	Sequence 7, Appl1
58	35	34.7	625	4 US-09-019-809-9	Sequence 9, Appl1
59	35	34.7	1157	2 US-08-532-547-5	Sequence 5, Appl1
60	35	34.7	1157	2 US-08-379-656B-5	Sequence 5, Appl1

## ALIGNMENTS

```
RESULT 1
5516630-2
; PATENT NO. 5516630
; APPLICANT: TICEHURST, JOHN R.;BALTIMORE, DAVID;FEINSTONE,
; STEPHEN M.;PURCELL, ROBERT H.;RACANIELLO, VINCENT R.;
; BAROUDY, BAHIGE M.
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/78,262
; FILING DATE: 06-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 256,135
; FILING DATE: 06-OCT-1988
; APPLICATION NUMBER: 654,942
; FILING DATE: 27-SEP-1984
; APPLICATION NUMBER: 537,911
; FILING DATE: 30-SEP-1983
; SEQ ID NO.: 2
; LENGTH: 1091
5516630-2

Query Match      100.0%; Score 101; DB 6; Length 1091;
Best Local Similarity 100.0%; Pred. No. 6,9e+09;
Matches 20; Conservative 0; Mismatches 0; Indels 0;

OY 1 FSOAKISLFYTEHEIMKFS 20
Db 1071 FSOAKISLFYTEHEIMKFS 1090

RESULT 2
US-08-475-886-2
; Sequence 2, Application US/08475886A
```

```

; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

```

```

Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 FSOAKISLFYTEHEHIMKFS 20
        |||
Db      834 FSOAKISLFYTEHEHIMKFS 853

```

```

RESULT      3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

```

```

Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 FSOAKISLFYTEHEHIMKFS 20
        |||
Db      834 FSOAKISLFYTEHEHIMKFS 853

```

```

RESULT      4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W

```

```

; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

```

```

Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 FSOAKISLFYTEHEHIMKFS 20
        |||
Db      834 FSOAKISLFYTEHEHIMKFS 853

```

```

RESULT      5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 202642620S1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

```

```

Query Match          100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 FSOAKISLFYTEHEHIMKFS 20
        |||
Db      834 FSOAKISLFYTEHEHIMKFS 853

```

```

RESULT      6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC

```

;; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
;; FILE REFERENCE: 202642620S1  
;; CURRENT APPLICATION NUMBER: US/08/397,232A  
;; CURRENT FILING DATE: 1995-04-17  
;; EARLIER APPLICATION NUMBER: 07/947,338  
;; EARLIER FILING DATE: 1992-09-18  
;; EARLIER APPLICATION NUMBER: PCT/US93/08610  
;; EARLIER FILING DATE: 1993-09-17  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO: 4  
;; LENGTH: 2227  
;; TYPE: PRT  
;; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1,5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEEHIMKFS 20  
|||  
Db 834 FSOAKISLFYTEEHIMKFS 853

RESULT 7  
US-09-171-387-2  
; Sequence 2, Application US/09171387  
; Patent No. 6280734  
; GENERAL INFORMATION:  
; APPLICANT: RAYCHAUDHURI, GOPA;  
; EMMERSON, SUZANNE, U.;  
; PURCELL, ROBERT, H.;  
; TITLE OF INVENTION: SIMIAN-HUMAN HAV  
; HAVING A CHIMERIC 2C PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MICROSOFT WORD 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,387  
; FILING DATE: 24-Mar-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/06506  
; FILING DATE: 18-APR-1997  
; APPLICATION NUMBER: US60/015,642  
; FILING DATE: 19-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: William S. Feller  
; REGISTRATION NUMBER: 26,728  
; REFERENCE/DOCKET NUMBER: 2026-42290S1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 2  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2227 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1,5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEEHIMKFS 20  
|||  
Db 834 FSOAKISLFYTEEHIMKFS 853

RESULT 8  
US-07-920-281C-2  
; Sequence 2, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920,281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-920-281C-2

Query Match 40.6%; Score 41; DB 1; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 2e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEEH 14  
|||  
Db 1072 FSAKVSLEYENN 1085

RESULT 9  
US-08-466-277-2  
; Sequence 2, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
US-08-466-277-2

ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,277  
FILING DATE: 06-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/920,281  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 828-103P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-241-1300  
TELEFAX: 703-241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2431 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-466-277-2

Query Match 40.6%; Score 41; DB 4; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 2e+02; 5; Indels 0; Gaps 0;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSOAKISLFTYEEH 14  
|||:|:|:|  
Db 1072 FSAKVSLEYENH 1085

RESULT 10  
PCT-US92-00282-11  
Sequence 11, Application PC/TUS9200282  
GENERAL INFORMATION:  
APPLICANT: OWENS, IDA S.  
APPLICANT: RITTER, JOSEPH K.  
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
TITLE OF INVENTION: THEREIN.  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
STREET: 1615 L STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/00282  
FILING DATE: 19920110  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US92-00282-11

Query Match 38.6%; Score 39; DB 5; Length 289;  
Best Local Similarity 58.3%; Pred. No. 41;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 FYTEHEIMKFS 20  
|||:|:|:|  
Db 100 FFTEHLMKFS 111

RESULT 11  
US-08-092-817-4  
Sequence 4, Application US/08092817  
Patent No. 5496926  
GENERAL INFORMATION:  
APPLICANT: RUBINSTEIN, Menachem  
APPLICANT: NOVICK, Daniela  
APPLICANT: TAL, Nathan  
TITLE OF INVENTION: SOLUBLE LDL RECEPTOR, ITS PRODUCTION AND  
TITLE OF INVENTION: USE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/092,817  
FILING DATE: 19-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 100696  
FILING DATE: 19-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 102915  
FILING DATE: 23-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: TOWNSEND, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: RUBINSTEIN-5A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 860 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-092-817-4



Query Match 38.68; Score 39; DB 1; Length 860;  
Best Local Similarity 46.28; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 LFYTEEHIMKFS 20  
11:111:1:  
DB 401 LEFTNRHEVRKMT 413

## RESULT 12

US-08-470-058-4  
Sequence 4, Application US/08470058  
Patent No. 5817789  
GENERAL INFORMATION:  
APPLICANT: Heartlein, Michael W.  
TITLE OF INVENTION: Chimeric Proteins For Use In Transport  
TITLE OF INVENTION: Of A Selected Substance Into Cells  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,058  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: TKT93-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1410 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-470-058-4

Query Match 38.68; Score 39; DB 2; Length 1410;  
Best Local Similarity 46.28; Pred. No. 2.3e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 LFYTEEHIMKFS 20  
11:111:1:  
DB 401 LEFTNRHEVRKMT 413

## RESULT 13

US-09-037-188-4  
Sequence 4, Application US/09037188  
Patent No. 6027921  
GENERAL INFORMATION:  
APPLICANT: Heartlein, Michael W.  
TITLE OF INVENTION: Chimeric Proteins For Use  
TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS  
NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/037,188  
FILING DATE: 02-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Ph.D., J.D., Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 07236/009002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1410 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-037-188-4

Query Match 38.68; Score 39; DB 3; Length 1410;  
Best Local Similarity 46.28; Pred. No. 2.3e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 LFYTEEHIMKFS 20  
11:111:1:  
DB 401 LEFTNRHEVRKMT 413

## RESULT 14

US-09-285-310-4  
Sequence 4, Application US/09285310  
Patent No. 6262026  
GENERAL INFORMATION:  
APPLICANT: Heartlein, Michael W.  
TITLE OF INVENTION: Chimeric Proteins For Use  
TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/285,310  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/037,188

ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Ph.D., J.D., Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 07236/009002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200134  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1410 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-285-310-4

Query Match 38.6%; Score 39; DB 4; Length 1410;  
Best Local Similarity 46.2%; Pred. No. 2.3e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 LFVTEHEIMKFS 20  
11:1 11:1  
Db 401 LFTYRHEVRKMT 413

RESULT 15  
US-07-958-551-2  
Sequence 2, Application US/07958551  
Patent No. 5302387  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M.  
APPLICANT: Kennedy, Keith M.  
APPLICANT: Randall, John Brooks  
APPLICANT: Brower, David Orlin  
TITLE OF INVENTION: Bacillus thuringiensis Isolates Active Against  
TITLE OF INVENTION: Cockroaches and Genes Encoding Cockroach-Active  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/958,551  
FILING DATE: October 19, 1992  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/788,654  
FILING DATE: No. 5302387ember 6, 1991  
NAME:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/J 101.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neolecensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: Lambdagem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-07-958-551-2

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KISFYTEHEI 16  
11:1 11:1  
Db 2 KESIVNEENI 13

RESULT 16  
US-08-129-610-7  
Sequence 7, Application US/08129610  
Patent No. 5436002  
GENERAL INFORMATION:  
APPLICANT: Jewel Payne  
APPLICANT: Kenneth Narva  
APPLICANT: Kendrick Akira Uyeda  
APPLICANT: Christine Julie Stalder  
TITLE OF INVENTION: No. 5436002el Bacillus thuringiensis Isolates and Toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/129,610  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,199  
FILING DATE: 15-JUL-1993  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/977,350  
FILING DATE: 17-NOV-1992  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/746,751  
FILING DATE: 21-AUG-1991  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/708,266  
FILING DATE: 28-MAY-1991  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,399  
FILING DATE: 29-JAN-1991  
NAME:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA55CCD.C1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: *Bacillus thuringiensis*  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS20176  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TW)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
J5-08-129-610-7

Query Match	37.6%	Score 38;	DB 1;	Length 265;
Best Local Similarity	58.3%	Pred. NO. 55;		
Matches 7;	Conservative	3;	Mismatches 2;	Indels 0;
				Gaps 0;

```

OY      5 KISLFYTEEHEI 16
         | | : | | | : |
Db      2 KESIYYNEENEI 13

```

RESULT 17  
 US-08-129-609A-7  
 Sequence 7, Application US/08129609A  
 Patent No. 5489432  
 GENERAL INFORMATION:  
 APPLICANT: Jewel Payne  
 APPLICANT: M. Keith Kennedy  
 APPLICANT: John Brookes Randall  
 APPLICANT: David Orlin Brower  
 APPLICANT: H. Ernest Schepf  
 TITLE OF INVENTION: *Bacillus thuringiensis* Isolates Active  
 TITLE OF INVENTION: Against Cockroaches and Genes Encoding Cockroach-Active Toxins  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: David R. Saliwanchik  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/129,609A  
 FILING DATE: 30-SEP-1993  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/958,551  
 FILING DATE: 19-OCT-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/788,654  
 FILING DATE: 6-NOV-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Saliwanchik, David R.  
 REGISTRATION NUMBER: 31,794  
 REFERENCE/DOCKET NUMBER: M/J 101.C2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-375-8100  
 TELEFAX: 904-372-5800

1 INFORMATION FOR SEQ ID NO: 7 :  
 2 SEQUENCE CHARACTERISTICS:  
 3 LENGTH: 265 amino acids  
 4 TYPE: amino acid  
 5 STRANDEDNESS: single  
 6 TOPOLOGY: linear  
 7 MOLECULE TYPE: protein  
 8 HYPOTHETICAL: YES  
 9 ANTI-SENSE: NO  
 10 ORIGINAL SOURCE:  
 11 ORGANISM: *Bacillus thuringiensis*  
 12 STRAIN: neolecensis  
 13 INDIVIDUAL ISOLATE: PS201t6  
 14 IMMEDIATE SOURCE:  
 15 LIBRARY: lambdaGem (TM)-11 Library of Kenneth E. Narva  
 16 CLONE: 201T635  
 17 US-08-129-609A-7

Query Match	37.6%	Score 38	DB 1	Length 265
Best Local Similarity	58.3%	Pred. NO. 55		
Matches 7, Conservative	3	Mismatches 2	Indels 0	Gaps 0

```
QY      5 KISLFYTEEHEI 16
        | | : : | | : |
Db      2 KESIYYNEENEI 13
```

```

1 RESULT 18
2 US-08-455-313-7
3 Sequence 7, Application US/08455313
4 Patent No. 5635480
5 GENERAL INFORMATION:
6 APPLICANT: Jewel Payne
7 APPLICANT: Kenneth Narva
8 APPLICANT: Kendrick Akira Uyeda
9 APPLICANT: Christine Julie Stalder
10 APPLICANT: Tracy Ellis Michaels
11 TITLE OF INVENTION: No. 5635480el Bacillus thuringiensis Isolates and Toxins
12 NUMBER OF SEQUENCES: 10
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: David R. Sallwanchik
15 STREET: 2421 N.W. 41st Street, Suite A-1
16 CITY: Gainesville
17 STATE: FL
18 COUNTRY: USA
19 ZIP: 32606
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: Patentln Release #1.0, Version #1.25
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/08/455,313
27 FILING DATE: 31-MAY-1995
28 CLASSIFICATION: 514
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 08/129,610
31 FILING DATE: 30-SEP-1993
32 APPLICATION NUMBER: US 08/093,199
33 FILING DATE: 15-JUL-1993
34 CLASSIFICATION: 514
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 07/977,350
37 FILING DATE: 17-NOV-1992
38 CLASSIFICATION: 514
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 07/746,751
41 FILING DATE: 21-AUG-1991
42 CLASSIFICATION: 514
43 PRIOR APPLICATION DATA:
44 APPLICATION NUMBER: US 07/708,266
45 FILING DATE: 28-MAY-1991

```

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,399  
FILING DATE: 29-JAN-1991  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA55CCD.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS20176  
IMMEDIATE SOURCE:  
LIBRARY: Lambdagem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-08-455-313-7

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 KISLFYEEHEI 16  
| : : | : : | : : |  
Db 2 KESIVYNEHEI 13

RESULT 19  
US-08-475-924-2  
Sequence 2, Application US/08475924  
Patent No. 5723440  
GENERAL INFORMATION:  
APPLICANT: Stockhoff, Brian  
APPLICANT: Conlan, Christopher  
TITLE OF INVENTION: CONTROLLING HEMIPTERAN INSECT PESTS  
TITLE OF INVENTION: WITH BACILLUS THURINGIENSIS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,924  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS20176  
IMMEDIATE SOURCE:  
LIBRARY: Lambdagem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-08-475-924-2

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 KISLFYEEHEI 16  
| : : | : : | : : |  
Db 2 KESIVYNEHEI 13

RESULT 20  
US-08-657-579A-2  
Sequence 2, Application US/08657579A  
Patent No. 5885963  
GENERAL INFORMATION:  
APPLICANT: Stockhoff, Brian  
APPLICANT: Conlan, Christopher  
TITLE OF INVENTION: CONTROLLING HEMIPTERAN INSECT PESTS  
TITLE OF INVENTION: WITH BACILLUS THURINGIENSIS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,579A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/475,924  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA96.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 352-375-8100  
TELEFAX: 352-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleogensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-08-657-579A-2

Query Match 37.6%; Score 38; DB 2; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 5 KISLYTEPHEI 16  
| : : : | : : : |  
Db 2 KESIIYNEBEI 13

## RESULT 21

US-09-224-025-7  
Sequence 7, Application US/09224025

GENERAL INFORMATION:  
PATENT No. 6150165  
APPLICANT: Jewel Payne  
APPLICANT: Kenneth Narva  
APPLICANT: Kendrick Akira Uyeda  
APPLICANT: Christine Julie Stalder  
APPLICANT: Tracy Ellis Michaels  
TITLE OF INVENTION: No. 6150165el Bacillus thuringiensis isolates and toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/224,025  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/455,313  
FILING DATE:  
APPLICATION NUMBER: US 08/093,199  
FILING DATE: 15-JUL-1993

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/977,350  
FILING DATE: 17-NOV-1992

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/746,751  
FILING DATE: 21-AUG-1991

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/708,266  
FILING DATE: 28-MAY-1991

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,399  
FILING DATE: 29-JAN-1991

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: MA55CCD.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleogensis  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-09-224-025-7

Query Match 37.6%; Score 38; DB 4; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 5 KISLYTEPHEI 16  
| : : : | : : : |  
Db 2 KESIIYNEBEI 13

## RESULT 22

PCT-US94-07887-7  
Sequence 7, Application PC/TUS9407887

GENERAL INFORMATION:  
APPLICANT: Street address: 4980 Carroll Canyon Road  
APPLICANT: City: San Diego  
APPLICANT: State/Province: California  
APPLICANT: Country: US  
APPLICANT: Postal code/Zip: 92121  
APPLICANT: Phone number: (619) 453-8030 Fax number: (619)453-6991  
APPLICANT: Telex number:  
TITLE OF INVENTION: Novel Bacillus thuringiensis isolates and toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07887  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,199  
FILING DATE: 15-JUL-1993

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/977,350  
FILING DATE: 17-NOV-1992

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/746,751  
FILING DATE: 21-AUG-1991

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/708,266  
FILING DATE: 28-MAY-1991  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,399  
FILING DATE: 29-JAN-1991  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA55CCD.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-372-5800  
TELEFAX: 904-375-8100  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: lambdagem (TM)-11 Library of Kenneth E. Narva  
PCT-US94-0787-7

Query Match 37.6%; Score 38; DB 5; Length 265;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KISLFYTEHEHET 16  
| : : | | : | |  
Db 2 KESIYNEENET 13

RESULT 23  
US-08-772-270A-4  
Sequence 4, Application US/08772270A  
Patent No. 6019984  
GENERAL INFORMATION:  
APPLICANT: MacInnes, Janet  
APPLICANT: Ricciardi, Paul  
APPLICANT: Mallard, Bonnie  
APPLICANT: Rosendal, Soren  
TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR  
TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/772,270A  
FILING DATE: December 23, 1996  
CLASSIFICATION: 42A  
ATTORNEY/AGENT INFORMATION:

NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 6580-81  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 707 amino acids.  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Actinobacillus pleuropneumoniae  
US-08-772-270A-4

Query Match 37.6%; Score 38; DB 3; Length 707;  
Best Local Similarity 44.4%; Pred. No. 1.6e+02;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEHET 18  
| : : | | : | |  
Db 629 FDEATSAIDYSEHET 646

RESULT 24  
US-08-097-827-2  
Sequence 2, Application US/08097827  
GENERAL INFORMATION:  
APPLICANT: Baum, Peter  
Goodwin, Ray  
Fanslow, William  
Gayle, Richard  
TITLE OF INVENTION: Novel Cytokine which is a ligand for  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,827  
FILING DATE: 23-Jul-1993  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia A.  
REGISTRATION NUMBER: 34,693  
REFERENCE/DOCKET NUMBER: 2806  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-587-0730  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 198 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-097-827-2

Query Match 36.6%; Score 37; DB 1; Length 198;  
Best Local Similarity 42.9%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEEH 14  
| : | | : | : |  
Db 111 FOEVKIDLHFREDH 124

RESULT 25  
US-08-494-574-2  
; Sequence 2, Application US/08494574  
; Patent No. 5783665  
; GENERAL INFORMATION:  
; APPLICANT: Baum, Peter  
; APPLICANT: Goodwin, Ray  
; APPLICANT: Fanslow, William  
; APPLICANT: Gayle, Richard  
; TITLE OF INVENTION: No. 578365el Cytokine which is a ligand for  
; TITLE OF INVENTION: OX40  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/494,574  
; FILING DATE: 22-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/097,827  
; FILING DATE: 23-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia A.  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 2806  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-587-0730  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 198 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-494-574-2

Query Match 36.6%; Score 37; DB 1; Length 198;  
Best Local Similarity 42.9%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
OY 1 FSOAKISLFYTEEH 14  
| : | | : | : |  
Db 111 FOEVKIDLHFREDH 124

Search completed: June 16, 2002, 00:03:17  
Job time: 8531 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:19 ; Search time 108.75 Seconds  
(without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-432a-43

Perfect score: 101  
Sequence: 1 FSOAKISLFYTEHEIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	1358	2 A03905	genome polypotein
2	101	100.0	2227	1 GNNYHM	genome polypotein
3	101	100.0	2227	1 GNNYHR	genome polypotein
4	101	100.0	2227	1 GNNYMK	genome polypotein
5	101	100.0	2227	1 GNNYHB	genome polypotein
6	97	96.0	852	1 GNNYHA	genome polypotein
7	88	87.1	2230	1 GNNYSA	genome polypotein
8	47	46.5	378	2 F64300	formate dehydrogen
9	43	42.6	325	2 H71869	hypothetical prote
10	43	42.6	410	2 T22413	hypothetical prote
11	42	41.6	125	2 AB1747	glycine cleavage s
12	42	41.6	125	2 A11377	glycine cleavage s
13	42	41.6	187	2 T18844	hypothetical prote
14	42	41.6	545	2 S10842	gene ND5 intron 2
15	42	41.6	1265	2 F88886	protein F5262.2 (1
16	41.5	41.1	288	2 S64243	hypothetical prote
17	41	40.6	144	2 E75128	glycine cleavage s
18	41	40.6	190	2 C90529	nuclease, 11poprot
19	41	40.6	209	2 S65300	hypothetical prote
20	41	40.6	304	2 C72324	homoserine O-succi
21	41	40.6	465	2 B97235	deacetylase/dipep
22	41	40.6	483	2 G84113	hypothetical prote
23	41	40.6	557	2 S58164	glucose-6-phosphat
24	41	40.6	820	2 T51510	hypothetical prote
25	41	40.6	1040	1 A38306	alpha-mannosidase
26	41	40.6	2431	1 MNMVSF	nonstructural poly
27	40.5	40.1	251	2 T13104	minor tail protein
28	40.5	40.1	419	2 T23577	hypothetical prote
29	40	39.6	142	2 C70059	hypothetical prote

30	40	39.6	203	2 S63633	hypothetical prote
31	40	39.6	256	2 T45895	hypothetical prote
32	40	39.6	298	2 A89009	protein T2764.1 (1
33	40	39.6	333	2 T33866	hypothetical prote
34	40	39.6	450	2 C82204	heat shock protein
35	40	39.6	523	2 D83631	probable sulfate t
36	39.5	39.1	109	2 S74054	hypothetical prote
37	39.5	39.1	260	2 B71353	probable D,D-carbo
38	39.5	39.1	410	2 T23264	hypothetical prote
39	39.5	39.1	507	2 B89792	hypothetical prote
40	39	38.6	129	2 F72075	hypothetical prote
41	39	38.6	129	2 G66547	hypothetical prote
42	39	38.6	145	2 F83247	hypothetical prote
43	39	38.6	152	2 C69202	conserved hypobet
44	39	38.6	167	2 F83845	hypothetical prote
45	39	38.6	171	2 T18167	hypothetical prote
46	39	38.6	289	2 B42586	glucuronosyltransf
47	39	38.6	306	1 BVBKR6	MK16 protein - ye
48	39	38.6	319	2 F81319	ADP-heptose--Lps h
49	39	38.6	350	2 AD1749	interalrin protein
50	39	38.6	380	1 NPBY	anthranilate phosp
51	39	38.6	382	2 E85082	hypothetical prote
52	39	38.6	382	2 T14186	hypothetical prote
53	39	38.6	398	2 F70353	RNA polymerase sig
54	39	38.6	433	2 A86535	hypothetical prote
55	39	38.6	433	2 A72099	hypothetical prote
56	39	38.6	435	2 A81580	hypothetical prote
57	39	38.6	581	2 B72221	hypothetical prote
58	39	38.6	645	2 D85680	unknown protein en
59	39	38.6	645	2 D85844	unknown protein en
60	39	38.6	645	2 A85629	major head protein

## ALIGNMENTS

RESULT 1  
A03905  
genome polypotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; prote  
C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
R:Baronuy, B.M.; Ticehurst, J.R.; Mele, T.A.; Malzel Jr., J.V.; Purcell, R.H.; Felns  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and R  
A:Reference number: A03905; MUID:8516289  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polypotein  
C:Keywords: coat protein; core protein; polypotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 101; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||  
DB 834 FSOAKISLFYTEHEIMKFS 853

RESULT 2  
GNNYHM  
genome polypotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J:Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with d1  
A:Reference number: A25981; M01D:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP2>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

Db 834 FSOAKISLFYTEHEIMKFS 853

RESULT 3  
GENYHR  
genome polypeptide - human hepatitis A virus  
N:contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Naajarian, R.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; M01D:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAU>  
A:Cross-references: GB:K02990; NID:g329596; PIDN:AAA5472.1; PID:g329597  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

Db 834 FSOAKISLFYTEHEIMKFS 853

RESULT 4  
GENYHR  
genome polypeptide - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison  
A:Reference number: A94149; M01D:8715701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA45471.1; PID:g329595  
A:Note: submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

Db 834 FSOAKISLFYTEHEIMKFS 853

RESULT 5  
GENYHR  
genome polypeptide - human hepatitis A virus (strain MBV)  
N:contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
Vpg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wessel, T.; Klein, R.; Wimmer, E.; Deinhard  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso  
A:Reference number: J50303; M01D:88045071  
A:Accession: J50303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP3>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein Vpg #status predicted <VPG>

F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKFS 20  
DB 834 FSOAKISLFYTEHEIMKFS 853

## RESULT 6

genome polyprotein - human hepatitis A virus (strain CR326) (fragment)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904  
R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.W.  
J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904; MUID:85185648  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <1IN>  
A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 96.0%; Score 97; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 4.8e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKFS 19  
DB 834 FSOAKISLFYTEHEIMKFS 852

## RESULT 7

genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tzarev, S.A.  
submitted to JIPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BAA00766.1; PID:g222598  
R:Tzarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure an  
A:Reference number: J01080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjapardze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhiko  
Submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:g61971; PIDN:CAA3490.1; PID:g930268

R:Balayan, M.S.; Kusov, Y.Y.; Andjapardze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizh  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian  
A:Reference number: S03965; MUID:89232168

A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461

C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 87.1%; Score 88; DB 1; Length 2230;  
Best Local Similarity 85.0%; Pred. No. 4.6e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKFS 20  
DB 838 FSOAKISLFYTEHEIMKFS 857

## RESULT 8

formate dehydrogenase (EC 1.2.1.2) alpha chain - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: F64300  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kalne, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc  
A:Reference number: A64300; MUID:96337999  
A:Accession: F64300  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-378 <BDU>  
A:Cross-references: GB:U67459; GB:L77117; NID:g2826236; PIDN:AAB97987.1; PID:g1498763  
C:Genetics:  
A:Map position: REV8474-7338  
C:Keywords: oxidoreductase

Query Match 46.5%; Score 47; DB 2; Length 378;  
Best Local Similarity 47.1%; Pred. No. 4.7;  
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 4 AKISLFTTEHEIMKFS 20  
DB 174 SKVTFTFTEHEIMKFS 190

## RESULT 9

hypothetical protein jhp0940 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 04-Mar-2000  
C:Accession: H71869  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999  
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557  
A:Accession: H71869  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-325 <ARN>  
A:Cross-references: GB:AE001523; GB:AE001439; NID:g4155513; PIDN:AAD06516.1; PID:g415551  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0940  
C:Superfamily: Helicobacter pylori hypothetical protein jhp0940

Query Match 42.6%; Score 43; DB 2; Length 325;  
Best Local Similarity 60.0%; Pred. No. 19;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 KISLFTTEHEIMKF 19  
| | | | | | | : | |  
Db 24 KISLFYNELMYVKF 38

RESULT 10  
T22413  
hypothetical protein F49C12.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22413  
R:Gardner, A.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19562  
A:Accession: T22413  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-410 <MIL>  
A:Cross-references: EMBL:268277; PIDN:CAA92512.1; GSPDB:GN00022; CESP:F49C12.8  
A:Experimental source: clone F49C12  
C:Genetics:  
A:Gene: CESP:F49C12.8  
A:Map position: 4  
A:Introns: 21/3; 302/3

Query Match 42.6%; Score 43; DB 2; Length 410;  
Best Local Similarity 36.8%; Pred. No. 24;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FSOAKISLFTTEHEIMKF 19  
| : : | | : : | |  
Db 156 FAMIRVGLEFLDHLINKF 174

RESULT 11  
AB1747  
glycine cleavage system protein H homolog lin2519 [imported] - Listeria innocua (strain  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AB1747  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maltournam, A.; Ma  
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A>Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1747  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-125 <GLA>  
A:Cross-references: GB:AC92022; PIDN:CA97746.1; PID:g16415041; GSPDB:GN00178  
A:Experimental source: strain Clp11262

C:Genetics:  
A:Gene: lin2519  
C:Superfamily: glycine cleavage system protein H; lipoY1/biotin-binding homology

Query Match 41.6%; Score 42; DB 2; Length 125;  
Best Local Similarity 72.7%; Pred. No. 9.3;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 LFYTEHEIMK 18  
| | | | | | : |  
Db 7 LFTYEEHWVK 17

RESULT 12  
A11377  
glycine cleavage system protein H homolog lmo2425 [imported] - Listeria monocytogenes  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: A11377  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maltournam, A.;  
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A>Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: A11377  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-125 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00503.1; PID:g16411913; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2425  
C:Superfamily: glycine cleavage system protein H; lipoY1/biotin-binding homology

Query Match 41.6%; Score 42; DB 2; Length 125;  
Best Local Similarity 72.7%; Pred. No. 9.3;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 LFYTEHEIMK 18  
| | | | | | : |  
Db 7 LFTYEEHWVK 17

RESULT 13  
T18844  
hypothetical protein C01H6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18844  
R:Berk, M.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19030  
A:Accession: T18844  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1187 <MIL>  
A:Cross-references: EMBL:271258; PIDN:CAA95778.1; GSPDB:GN00019; CESP:C01H6.3  
A:Experimental source: clone C01H6  
C:Genetics:  
A:Gene: CESP:C01H6.3  
A:Map position: 1  
A:Introns: 42/3; 82/1; 109/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein C01H6.3

Query Match 41.6%; Score 42; DB 2; Length 187;  
Best Local Similarity 36.8%; Pred. No. 15;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;













```
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PMW: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT
CC SHOWN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M14114; AAA45475.1; -.
CC EMBL: M14707; AAA45465.1; -.
CC EMBL: M14707; AAA45466.1; ALT_INIT.
CC EMBL: M16632; AAA45471.1; -.
CC PIR: A25981; GNNYHM.
CC PIR: A25914; GNNYMK.
CC PIR: A03905; A03905.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol_1.
CC Pfam: PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC -----
CC CHAIN 1 23
CC CHAIN 24 245
CC CHAIN 246 491
CC CHAIN 492 836
CC CHAIN 837 980
CC CHAIN 981 1087
CC CHAIN 1088 1422
CC CHAIN 1423 1496
CC CHAIN 1497 1519
CC CHAIN 1520 1738
CC CHAIN 1739 2227
CC CHAIN 77 77
CC CHAIN 764 764
CC CHAIN 821 821
CC CHAIN 1052 1052
CC CHAIN 1062 1062
CC CHAIN 1118 1118
CC CHAIN 1151 1151
CC CHAIN 1163 1163
CC CHAIN 1277 1277
CC CHAIN 1500 1500
CC CHAIN 1805 1805
CC CHAIN 1930 1930
CC CHAIN 2227 AA; 251506 MW; 01E225E7AB740A6 CRC64;
CC SEQUENCE
CC -----
Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain LA).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=65190549; PubMed=2986127;
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA Merryweather J., van Nest G., Dina D.,
RT Primary structure and gene organization of human hepatitis A virus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PMW: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K02990; AAA45472.1; -.
CC PIR: A03903; GNNYHR.
CC MEROPS: C03.005; -.
CC InterPro: IPR000605; RNA_helicase.
CC InterPro: IPR001205; RNA_pol_P3D.
CC Pfam: PF00680; RNA_dep_RNA_pol_1.
CC Pfam: PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC -----
CC CHAIN 1 23
CC CHAIN 24 245
CC CHAIN 246 491
CC CHAIN 492 836
CC CHAIN 837 980
CC CHAIN 981 1076
CC CHAIN 1077 1422
CC CHAIN 1423 1484
CC CHAIN 1485 1507
CC CHAIN 1508 1678
CC CHAIN 1679 2227
CC CHAIN 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;
CC SEQUENCE
CC -----
Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCB1\_TaxID=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86045071; PubMed=2623500;  
RA Deinhardt F.;  
RA Paul A.V., Tada H., der Helm K., Wiesel T., Klein R., Wimmer E.,  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
virus (isolate MBB)."  
RL Virus Res. 8:153-171(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M20273; AAA5474.1; -  
DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol\_1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT CHAIN 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
SO SEQUENCE

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKFS 20  
DB 834 FSOAKISLFYTEHEIMKFS 853

RESULT 4  
POLG\_HPAVC STANDARD; PRT; 852 AA.  
AC P06442; Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
DE P2A] (Fragment).  
OS Hepatitis A virus (strain CR326).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCB1\_TaxID=12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85185648; PubMed=2985793;

RA Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
RA Young A., Mitra S.W.;  
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA."  
RL J. Virol. 54:247-255(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M10033; AAA5470.1; -  
DR PIR: A03904; GNNYHA.  
DR Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 >852  
FT CHAIN 852 852  
FT NON\_TER  
SO SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;

Query Match 96.0%; Score 97; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEIMKFS 19  
DB 834 FSOAKISLFYTEHEIMKFS 852

RESULT 5  
POLG\_HPAV2 STANDARD; PRT; 2226 AA.  
AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCB1\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstein S.M.,  
RA Cremons T., Jansen R.W., Shelds P.A., Ping L.H., Feinstein S.M.,  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M59810; AAA45468.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease;  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 92.1%; Score 93; DB 1; Length 2226;  
Best Local Similarity 90.0%; Pred. No. 2.1e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FSOAKISLFYEEHEMKFS 20  
DB 834 FSOAKISLFYEEHEMKFS 853  
|||||

RESULT 6  
POLG\_HPAV4 STANDARD; PRT; 2226 AA.  
AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 43c).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91162758; PubMed-1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstein S.M.,  
Crommans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
variants arising during persistent infection: evidence for genetic  
recombination."  
RT J. Virol. 65:2056-2065(1991).  
RL -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PMM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M59809; AAA45469.1; -  
DR MEROPS; C03.005; -

DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein;  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09B7F5 CRC64;

Query Match 92.1%; Score 93; DB 1; Length 2226;  
Best Local Similarity 90.0%; Pred. No. 2.1e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FSOAKISLFYEEHEMKFS 20  
DB 834 FSOAKISLFYEEHEMKFS 853  
|||||

RESULT 7  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91162758; PubMed-1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstein S.M.,  
Crommans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
variants arising during persistent infection: evidence for genetic  
recombination."  
RT J. Virol. 65:2056-2065(1991).  
RL -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
CC -1- PMM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M59808; AAA45467.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;

KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
DR CHAIN 24 245 COAT PROTEIN VP4 (PIA).  
FT CHAIN 246 491 COAT PROTEIN VP2 (PIB).  
FT CHAIN 492 794 COAT PROTEIN VP3 (PIC).  
FT CHAIN 795 900 COAT PROTEIN VP1 (PID).  
FT CHAIN 901 1087 CORE PROTEIN P2A.  
FT CHAIN 1088 1422 CORE PROTEIN P2B.  
FT CHAIN 1423 1495 CORE PROTEIN P2C.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3A.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3B.  
FT CHAIN 1738 2226 PROBABLE PROTEIN P3C.  
SO SEQUENCE 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;

Query Match 92.1%; Score 93; DB 1; Length 2226;  
Best Local Similarity 90.0%; Pred. No. 2.1e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEHMKFS 20  
DB 834 FSOAKISLFYTEHEHMKFS 853  
|||||  
POLG\_HPAVS STANDARD; PRT; 2230 AA.  
AC P14553:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE Simian hepatitis A virus (strain AGM-27).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311420; PubMed=1649901;  
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,  
RA Purcell R.H.;  
RT \*Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE=89232168; PubMed=2541023;  
RA Balayan M.S., Kusov Y.Y., Andjapardze A.G., Tsarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Bilnov V.M., Vasilenko S.K.;  
RT \*Variations in genome fragments coding for RNA polymerase in human and simian hepatitis A viruses.";  
RL PDBS left. 247:425-428(1989).  
CC -1 SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1 SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1 SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00924; BAA00766.1; -;  
DR EMBL; X15461; CAA33490.1; -;  
DR PIR; A30470; GNNYSA.  
DR PIR; S04885; S04885.  
DR MEROPS; C03.005; -;

DR Interpro; IPR000605; RNA\_helicase.  
DR Interpro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam; PF00910; RNA\_helicase\_1.  
KM Polypeptide; Coat protein; Core protein; Hydrolase; Transferase;  
FT CHAIN 1 27  
DR CHAIN 28 249 COAT PROTEIN VP4 (PIA).  
FT CHAIN 250 495 COAT PROTEIN VP2 (PIB).  
FT CHAIN 496 795 COAT PROTEIN VP3 (PIC).  
FT CHAIN 796 984 COAT PROTEIN VP1 (PID).  
FT CHAIN 985 1091 CORE PROTEIN P2A.  
FT CHAIN 1092 1426 CORE PROTEIN P2B.  
FT CHAIN 1427 1498 CORE PROTEIN P2C.  
FT CHAIN 1499 1521 PROBABLE PROTEIN 3A.  
FT CHAIN 1522 1741 PROBABLE PROTEIN 3B.  
FT CHAIN 1742 2230 PROBABLE PROTEIN 3C.  
SO SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match 87.1%; Score 88; DB 1; Length 2230;  
Best Local Similarity 85.0%; Pred. No. 1.4e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEHMKFS 20  
DB 838 FSOAKISLFYTEHEHMKFS 857  
|||||  
POLG\_METJA STANDARD; PRT; 378 AA.  
AC P06314:  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Putative formate dehydrogenase alpha chain (EC 1.2.1.2).  
DE Mj0006.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.D., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Uitterlbeck T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser K.M., Smith H.O., Woese C.R., Venter J.C.;  
RT \*Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1 CATALYTIC ACTIVITY: Formate + NAD(+) -> CO(2) + NADH.  
CC -1 COPACITOR: MOLYBDENUM (MOLYBDOPTERIN), ZINC, FAD. MAY BIND A 4FE-4S CLUSTER (BY SIMILARITY).  
CC -1 SUBUNIT: DIMER OF ALPHA AND BETA CHAINS (BY SIMILARITY).  
CC -1 SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTEIN-CONTAINING OXIDOREDUCTASE FAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U67459; AAB97987.1; -;

DR HSSP; P07658; 1AA6.  
 DR TIGR; M00006; -  
 DR InterPro; IPR001467; Molybdopterin.  
 DR Pfam; PF00384; molybdopterin.2.  
 DR PROSITE; PS00551; MOLYBDOPTERIN\_PROK\_1; 1.  
 DR PROSITE; PS00490; MOLYBDOPTERIN\_PROK\_2; 1.  
 DR PROSITE; PS00932; MOLYBDOPTERIN\_PROK\_3; FALSE\_NEG.  
 DR Hypothetical protein; Oxidoreductase; Zinc; Flavoprotein; Molybdenum;  
 KM FAD; NAD; Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 8 8 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 11 11 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 15 15 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 43 43 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 378 AA; 42050 MW; E0C13060F2609AA CRC64;

Query Match 46.5%; Score 47; DB 1; Length 378;  
 Best Local Similarity 47.1%; Pred. No. 1.7;  
 Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEHEIMKFS 20  
 : : : : :  
 Db 174 SKVITFTEKEELKLN 190

RESULT 10  
 YE28\_CAEL STANDARD; PRT; 410 AA.  
 ID YE28\_CAEL  
 AC Q20585;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 47.6 kDa protein F49G12.8 in chromosome IV.  
 GN F49G12.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Gardner A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE FUS6 FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z68227; CA92512.1; -  
 DR WormRep; F49G12.8; CE03368.  
 DR InterPro; IPR000717; PCT.  
 DR Pfam; PF01399; PCT; 1.  
 DR SMART; SM00088; PINT; 1.  
 KM Hypothetical protein.  
 SO SEQUENCE 410 AA; 47583 MW; F37390A3250109EE CRC64;

Query Match 42.6%; Score 43; DB 1; Length 410;  
 Best Local Similarity 36.8%; Pred. No. 8.7;  
 Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEHEIMK 19  
 : : : : :  
 Db 156 FAMIRVGLFELDHLINKF 174

RESULT 11  
 YC06\_KLEPN

ID YC06\_KLEPN STANDARD; PRT; 722 AA.  
 AC Q48452;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 80.4 kDa protein in CPS region (ORF6).  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CHEIDID;  
 RX MEDLINE=95204345; PubMed=7896702;  
 RA Arakawa Y., Wacharotayanankun R., Nagatauka T., Ito H., Kato N.,  
 RA Ohta M.;  
 RT "genomic organization of the Klebsiella pneumoniae cps region  
 RT responsible for serotype K2 capsular polysaccharide synthesis in the  
 RT virulent strain Cheidid.";  
 RL J. Bacteriol. 177:1788-1796(1995).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D12424; BAA0477.1; -  
 KM Hypothetical protein.  
 SO SEQUENCE 722 AA; 80400 MW; 3CAD6910AEB1C3D7 CRC64;

Query Match 44.4%; Score 43; DB 1; Length 722;  
 Best Local Similarity 44.4%; Pred. No. 16;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 FSOAKISLFYTEHEHEIMK 18  
 : : : : :  
 Db 326 FRESEISQVTKHEPTVK 343

RESULT 12  
 NIF3\_YEAST STANDARD; PRT; 288 AA.  
 ID NIF3\_YEAST  
 AC P53081;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NG1-interacting factor 3.  
 GN NIF3 OR YGL221C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE=97435481; PubMed=9290212;  
 RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;  
 RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae  
 RT chromosome VII.";  
 RL Yeast 13:1077-1090(1997).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC	EMBL; 272743; CA96937.1; -. SCD; S0003189; NTF3. DR InterPro; IPR02678; DUF34. Pfam; PF01784; DUF34.1. DR SEQUENCE 288 AA; 3188 MW; B6AB6E48AAEF4776A CRC64;
CC	EMBL; 272743; CA96937.1; -. SCD; S0003189; NTF3. DR InterPro; IPR02678; DUF34. Pfam; PF01784; DUF34.1. DR SEQUENCE 288 AA; 3188 MW; B6AB6E48AAEF4776A CRC64;

Query Match	41.18	Score	41.5	DB	1	Length	288
Best Local Similarity	41.78	Pred	No. 10				
Matches	10	Conservative	5	Mismatches	4	Indels	5
						Gaps	2

```
QY      1  FSOAK--ISLFYTEE---HEIMKF 19
          | | | : | | | | | | : |
Db      217 FKQLKEDVDLYTTGEMSHHEVLKW 240
```

RESULT 13

ID	GCSH_PYRAB	STANDARD;	PRT;	138 AA
1	1	1	1	1

AC Q9V06L;

DT 16-oct-2001 (Rel. 40, Created)  
DT 16-oct-2001 (Rel. 40, Last sequence update)  
DT 16-oct-2001 (Rel. 40, Last annotation update)  
DT 16-oct-2001 (Rel. 40, Last annotation update)  
DE Probable glycine cleavage system H protein.  
GN GCVIH OR PAB0559.  
OS *Pyrococcus abyssi*.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ORSA;  
RA Hellig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
RT structure and evolution.";  
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.

CC  
CC -1- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYZES THE DEGRADATION OF  
CC GLYCINE. THE H PROTEIN SHUTTLES THE METHYLAMINE GROUP OF GLYCINE  
CC FROM THE P PROTEIN TO THE T PROTEIN (BY SIMILARITY).  
CC -1- COPACOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COPACOR  
CC (BY SIMILARITY).  
CC -1- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:  
CC P, T, L, AND H (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE GCVH FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
CC .....  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as their content is in no way  
CC modified and this statement is not removed. Usage by, and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

```

DR  EMBL; AJ248285; CAB9742.1; ALT_INIT.
DR  HSPB; P16048; 1IMP.
DR  InterPro; IPR002830; GCY_H.
DR  InterPro; IPR003016; LipoYL.
DR  Pfam; PF01597; GCY_H; 1.
DR  PROSITE; PS00189; LipoYL; 1.
KW  LipoYL; Complete proteome.
EQ  BINDING 74 LipoYL (BY SIMILARITY).
FO  SEQUENCE 138 AA; 15545 MW; A015FC4B695A7592 CRC64

```

Query Match	40.6%	Score 41	DB 1	Length 138
Best Local Similarity	53.8%	Pred. No.	5.6	
Matches 7	Conservative 3	Mismatches 3	Indels 0	Gaps 0

```
Qy      3 QAKISLFYTEENE 15
          : | | | | |
Db     12 KVEGGLYTTKEHE 24
```

RESULT	14
META_THEME	
ID	META_THEME
STANDARD;	
PRT;	304 AA

AC Q9W2Y3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Homoserine O-succinyltransferase (EC 2.3.1.46) (Homoserine O-transuccinylase) (HTS).  
 DE

OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga  
NCBI\_TaxID=2336;

RN SEQUENCE FROM N.A.  
 RI [1]  
 RP STRAIN=MSB8 / DSM 3109;  
 RC MEDLINE=99287316; PubMed=10360571;  
 RX Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 RA Hat D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RL genome sequence of *Thermotoga maritima*.";  
 RI Nature 399:323-329(1999).

CC -1- CATALYTIC ACTIVITY: Succinyl-L-homoserine - COA + 0-  
succinyl-L-homoserine.  
CC -1- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTS VARIANT: FIRST STEP  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
CC -1- SIMILARITY: BELONGS TO THE HTS FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC	EMBL	AE001753	AAD35962.1	-	-
DR	TIGR	TM0881	-	-	-
KW	Methionine biosynthesis	Transferase	Acyltransferase		
FT	Complete proteome				
ACCESSION	142		POTENTIAL		
SEQUENCE	304 AA	35759 MW	3ED8226AABF8044E	CRC64	

Query Match	40.6%	Score 41	DB 1	Length 304
Best Local Similarity	36.0%	Pred. No. 14		
Matches	9	Conservative	4	Mismatches 8; Gaps 1;

```
QY      3 QAKISLFYTEEHE-----IMKF 19
          |::| || | : |::|
Db     64 QVNVTLLYETTHKPKHTPIEHILKF 88
```

ID	G6PI_ACICA	STANDARD;	PRT;	557	AA.
----	------------	-----------	------	-----	-----

AC Q59088; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose  
DE isomerase) (PGI) (Phosphohexose isomerase) (PHI).

GN Pg1.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae.  
OC Acinetobacter.  
OX NCBI\_TaxID=471;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-BD4;  
RA Stark M., Kaplan N., Ron E.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate - D-fructose 6-phosphate.  
CC -1- PATHWAY: INVOLVED IN GLYCOLYSIS AND IN GLUCONEOGENESIS.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: X89900; CAA61993.1; -  
DR InterPro: IPR001672; G6P\_Isomerase.  
DR Pfam: PF00342; PGI; 1.  
DR PRINTS: PR00662; G6PISOMERASE.  
DR PROSITE: PS00765; P\_GLUCOSE\_ISOMERASE\_1; 1.  
DR PROSITE: PS00174; P\_GLUCOSE\_ISOMERASE\_2; 1.  
KW Isomerase; Glucosyltransferase; Glycolysis.  
FT ACT\_SITE 392 392 BY SIMILARITY.  
FT ACT\_SITE 520 520 BY SIMILARITY.  
SQ SEQUENCE 557 AA; 63280 MW; D74AE214B139E4DC CRC64;  
  
OY 1 FSOAKSLFTEHEHMFES 20  
||| ||: |||: | | : :  
Db 87 FSOAKIN--YTEQRAMHWA 104  
  
Query Match 40.6%; Score 41; DB 1; Length 557;  
Best Local Similarity 50.0%; Pred. No. 26;  
Matches 10; Conservative 4; Mismatches 4; Indels 2; Gaps 1;  
  
RESULT 16  
MAN1\_RAT  
ID MAN1\_RAT STANDARD; PRT; 1040 AA.  
AC P21139;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Alpha-mannosidase (EC 3.2.1.24) (Alpha-D-mannoside manohydrolase)  
DE (AAMN).  
OS Rattus norvegicus (Rat).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Liver;  
RX MEDLINE-91009139; PubMed-2211613;  
RA Bischoff J., Moremen K., Lodish H.F.;  
RT "Isolation, characterization, and expression of cDNA encoding a rat  
RT liver endoplasmic reticulum alpha-mannosidase.";  
RL J. Biol. Chem. 265:17110-17117(1990).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-  
CC mannose residues in alpha-D-mannosides.  
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum.  
CC -1- PFM: A SOLUBLE ISOMER OF THE RAT LIVER MANNOSIDASE IS SUPPOSED TO  
CC BE DERIVED FROM ENOPLASMIC RETICULUM MEMBRANE ALPHA-MANNOSIDASE  
CC BY PROTEOLYSIS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: M57547; AAA41565.1; -  
DR PIR: A38306; A38306.  
DR InterPro: IPR000602; Glyco\_hydro\_38.  
DR Pfam: PF01074; Glyco\_hydro\_38; 2.  
KW Hydrolase; Glycosylase; Endoplasmic reticulum.  
SQ SEQUENCE 1040 AA; 115970 MW; 144F6B97C8F7EA7 CRC64;  
  
OY 2 SOAKSLFTEHEHMFES 17  
||| ||: |||: | | : :  
Db 173 SOAKLAVFHRDYNHNL 188  
  
Query Match 40.6%; Score 41; DB 1; Length 1040;  
Best Local Similarity 37.5%; Pred. No. 53;  
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
  
RESULT 17  
POLN\_SFV  
ID POLN\_SFV STANDARD; PRT; 2431 AA.  
AC P08411;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nonstructural polypeptide [Contains: Nonstructural proteins NSP1 TO  
DE NSP4].  
OS Semliki forest virus (SFV).  
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
CC Alphavirus.  
OX NCBI\_TaxID=11033;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-86286581; PubMed-3488539;  
RA Takkinen K.;  
RT "Complete nucleotide sequence of the nonstructural protein genes of  
RT Semliki Forest virus.";  
RL Nucleic Acids Res. 14:5667-5682(1986).  
CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.  
CC -1- PFM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL: X04129; CAA27741.1; -  
DR PIR: A23592; MNWVSF.  
DR MEROPS: C09.001; -  
DR InterPro: IPR002589; A1pp.  
DR InterPro: IPR002620; Peptidase\_C9.  
DR InterPro: IPR001788; RNA\_dep\_RNApol2.  
DR InterPro: IPR000606; Viral\_helicase1.  
DR Pfam: PF01661; A1pp; 1.  
DR Pfam: PF01707; Peptidase\_C9; 1.  
DR Pfam: PF00978; RNA\_dep\_RNApol2; 1.  
DR Pfam: PF01443; Viral\_helicase1; 1.  
DR SMART: SM00506; A1pp; 1.  
KW Polypeptide; Nonstructural protein; RNA-binding; Helicase.  
FT CHAIN 1 537 NONSTRUCTURAL PROTEIN NSP1.  
FT CHAIN 538 1335 NONSTRUCTURAL PROTEIN NSP2.  
FT CHAIN 1336 1817 NONSTRUCTURAL PROTEIN NSP3.  
FT CHAIN 1818 2431 NONSTRUCTURAL PROTEIN NSP4.  
SQ SEQUENCE 2431 AA; 269286 MW; 1F9EBA1022EE3EC5F CRC64;  
  
Query Match 40.6%; Score 41; DB 1; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 1,4e+02;







DR EMBL; M84124; AAA61247.1; ALT\_SEQ.  
DR EMBL; M84122; AAA61247.1; JOINED.  
DR EMBL; M84123; AAA61247.1; JOINED.  
DR MIM; 606430; -  
DR MIM; 191740; -  
DR InterPro; IPR002213; UDPGT.  
DR Pfam; PF00201; UDPGT; 1.  
DR PROSITE; PS00375; UDPGT; 1.  
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;  
KW Multigene family; Microsome; Alternative splicing.  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 534 UDP-GLUCURONOSYLTRANSFERASE 1-5.  
FT TRANSMEM 492 508 POTENTIAL.  
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 534 AA; 60071 MW; 05989F2A18EEFA4 CRC64;

Query Match 38.6%; Score 39; DB 1; Length 534;  
Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 FYTEHEIMKFS 20  
DB 100 FETEHLMKFS 111

RESULT 22  
YCX3\_EUGGR STANDARD; PRT; 634 AA.  
AC P31916; P31917;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 78.0 kDa protein in P5BC Intron 2 (ORF635).  
OS Euglena gracilis.  
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.  
OX NCBI\_TaxID=3039;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-Z.  
RX MEDLINE=93347989; Pubmed=8346031;  
RA Hallick R.B., Hong L., Dräger R.G., Favreau M.R., Monfort A.,  
RA Oresat B., Spielmann A., Stutz E.;  
RT "Complete sequence of Euglena gracilis chloroplast DNA.";  
RL Nucleic Acids Res. 21:3537-3544(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z11874; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; X70810; CAA50080.1; -  
DR PIR; S34499; S34499.  
DR PIR; S34500; S34500.  
KW Chloroplast; Hypothetical protein.  
SQ SEQUENCE 634 AA; 78049 MW; D966B8864519E334 CRC64;

Query Match 38.6%; Score 39; DB 1; Length 634;  
Best Local Similarity 63.6%; Pred. No. 67;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 FYTEHEIMKFS 19  
DB 353 FYTESFRILKF 363

RESULT 23  
HLV2\_ECOLI STANDARD; PRT; 707 AA.  
AC P10089;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemolysin secretion ATP-binding protein, chromosomal.  
GN HLVB  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-J96 / SEROTYPE O4;  
RX MEDLINE=85234404; Pubmed=3891743;  
RA Felmelee T., Pellet S., Welch R.A.;  
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";  
RL J. Bacteriol. 163:94-105(1985).  
CC -1- FUNCTION: INVOLVED IN THE EXPORT OF HEMOLYSIN A.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M10133; AAA23976.1; -  
DR PIR; B24433; LEECB.  
DR InterPro; IPR003593; AAA.  
DR InterPro; IPR001140; ABC\_transporter\_tmem.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR001687; ATP\_GMP\_A.  
DR Pfam; PF00664; ABC\_membrane\_1.  
DR Pfam; PF00005; ABC\_tran\_1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Hemolysis; Transport; ATP-binding; Transmembrane.  
FT TRANSMEM 158 179 POTENTIAL.  
FT TRANSMEM 187 204 POTENTIAL.  
FT TRANSMEM 269 289 POTENTIAL.  
FT TRANSMEM 293 311 POTENTIAL.  
FT TRANSMEM 360 377 POTENTIAL.  
FT TRANSMEM 387 403 POTENTIAL.  
FT TRANSMEM 407 425 POTENTIAL.  
FT NP BIND 502 509 ATP (BY SIMILARITY).  
SQ SEQUENCE 707 AA; 79463 MW; 21009CB45E59437E CRC64;

Query Match 38.6%; Score 39; DB 1; Length 707;  
Best Local Similarity 44.4%; Pred. No. 75;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 FSQAKISLFTYEHEIMK 18  
DB 629 FDEATSAIDYSEHVIHM 646

RESULT 24  
HLVB\_ECOLI STANDARD; PRT; 707 AA.  
AC P08716;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemolysin secretion ATP-binding protein, plasmid.



Search completed: June 16, 2002, 00:10:10  
Job time: 658 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:52 ; Search time 204.58 Seconds  
(Without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-43

Perfect score: 101  
Sequence: 1 FSOAKISLFTYEEHIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 60 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhcc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvivirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	112	12	086534 hepatitis a
2	101	100.0	258	12	099187 hepatitis a
3	101	100.0	1124	12	084780 hepatitis a
4	101	100.0	1161	12	005794 hepatitis a
5	101	100.0	2216	12	09WMA2 hepatitis a
6	101	100.0	2218	12	067824 hepatitis a
7	101	100.0	2218	12	067817 hepatitis a
8	101	100.0	2225	12	09D132 hepatitis a
9	101	100.0	2227	12	067825 hepatitis a
10	101	100.0	2227	12	067826 hepatitis a
11	101	100.0	2227	12	09WMA4 hepatitis a
12	101	100.0	2227	12	09WMA3 hepatitis a
13	101	100.0	2227	12	09WMA1 hepatitis a
14	101	100.0	2227	12	09WMA0 hepatitis a
15	101	100.0	2227	12	09WMA9 hepatitis a
16	98	97.0	184	12	087092 simian hepa

17	98	97.0	2227	12	091FH5 hepatitis a
18	86	85.1	251	12	09ENR1 hepatitis a
19	86	85.1	251	12	09ENR0 hepatitis a
20	86	85.1	251	12	09ENR9 hepatitis a
21	86	85.1	251	12	09ENR8 hepatitis a
22	86	85.1	251	12	09ENR7 hepatitis a
23	86	85.1	251	12	09ENR6 hepatitis a
24	86	85.1	251	12	09ENR5 hepatitis a
25	86	85.1	251	12	09ENR4 hepatitis a
26	86	85.1	251	12	09ENR3 hepatitis a
27	86	85.1	251	12	09ENR2 hepatitis a
28	86	85.1	251	12	09ENR1 hepatitis a
29	86	85.1	251	12	09ENR0 hepatitis a
30	86	85.1	251	12	09ENR9 hepatitis a
31	86	85.1	251	12	09ENR8 hepatitis a
32	86	85.1	251	12	09ENR7 hepatitis a
33	86	85.1	251	12	09ENR6 hepatitis a
34	86	85.1	251	12	09ENR5 hepatitis a
35	86	85.1	251	12	09ENR4 hepatitis a
36	86	85.1	251	12	09ENR3 hepatitis a
37	86	85.1	251	12	09ENR2 hepatitis a
38	86	85.1	251	12	09ENR1 hepatitis a
39	86	85.1	251	12	09ENR0 hepatitis a
40	86	85.1	251	12	09ENR9 hepatitis a
41	86	85.1	251	12	09ENR8 hepatitis a
42	86	85.1	251	12	09ENR7 hepatitis a
43	86	85.1	251	12	09ENR6 hepatitis a
44	86	85.1	251	12	09ENR5 hepatitis a
45	86	85.1	251	12	09ENR4 hepatitis a
46	77	76.2	126	12	09ENR3 hepatitis a
47	73	72.3	251	12	09ENR2 hepatitis a
48	72	71.3	125	12	09ENR1 hepatitis a
49	72	71.3	251	12	09ENR0 hepatitis a
50	49	48.5	94	12	09ENR9 hepatitis a
51	49	48.5	94	12	09ENR8 hepatitis a
52	49	48.5	94	12	09ENR7 hepatitis a
53	49	48.5	94	12	09ENR6 hepatitis a
54	49	48.5	94	12	09ENR5 hepatitis a
55	49	48.5	94	12	09ENR4 hepatitis a
56	49	48.5	94	12	09ENR3 hepatitis a
57	49	48.5	94	12	09ENR2 hepatitis a
58	49	48.5	94	12	09ENR1 hepatitis a
59	49	48.5	94	12	09ENR0 hepatitis a
60	49	48.5	94	12	09ENR9 hepatitis a

#### ALIGNMENTS

RESULT 1  
086534 . . . . . PRELIMINARY: PRT: 112 AA.  
AC 086534: . . . . .  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VPI/2A (FRAGMENT).  
GN VPI/2A.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCB1\_Taxid=12092;  
OX [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN=LSH/S.  
RX MEDLINE=92348853; Pubmed=1668326;  
RA Flueschi N., Cavallieri F., Garelick H., Prugnola A., Pellegrini V.,  
RT Zuckerman A.J.;  
RT "Characterization of a hepatitis A virus strain suitable for vaccine  
production".  
RL J. Hepatol. 13:S146-S151(1991).  
DR EMBL: S44105; AAB2739.2; -.  
FT NON\_TER 112 112

SO SEQUENCE 112 AA; 13145 MW; 3C0CBB84570D9A2C CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 112;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

DB 75 FSOAKISLFYTEHEIMKFS 94

RESULT 2  
ID 0991S7 PRELIMINARY; PRT; 258 AA.

AC 0991S7;  
DT 01-JUN-2001 (TEMBLrel. 17, Created)  
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L-A-1;  
RA Jiang C., Wng P., Zhang H., Liu J.;  
RT "Comparison of L-A-1 strain of Hepatitis A virus passaged differently  
in 2B5 cell.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AF333097; AKI19545.1; -.  
FT NON\_TER 1  
FT 258  
SO SEQUENCE 258 AA; 29292 MW; BB426B7312449327 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 258;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

DB 211 FSOAKISLFYTEHEIMKFS 230

RESULT 3  
ID 084780 PRELIMINARY; PRT; 1124 AA.

AC 084780;  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)  
(FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ochinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,  
Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prihod'ko G.G.,  
RA Blinov V.M., Vasilenko S.K., Sandakhchiev L.S., Kusov Y.Y.,  
RA Grabov V.I., Flier G.P., Balyan M.S., Drozdov S.G.;  
RL DOKL. Biochem. 285:379-383(1986).  
DR EMBL, X04200; CAA27297.1; -.  
EMBL, A11312; CAA00953.1; -.  
KW Nonstructural protein.  
FT NON\_TER 1  
FT 1124  
SO SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 1124;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

DB 791 FSOAKISLFYTEHEIMKFS 810

RESULT 4  
ID 005794 PRELIMINARY; PRT; 1161 AA.

AC 005794; Q67800; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
DE P3D (EC 2.7.7.48)) (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,  
RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,  
RA Balyan M.S.;  
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
DR EMBL, X15464; CAA33492.1; -.  
DR Interpro: IPR000408; RCCL  
KW PROSITE, PS00626; RCCL\_2; UNKNOWN\_1.  
KW Polypeptide; Coat protein; Core protein; RNA-directed RNA polymerase;  
KW Hydrolase; Thiol protease.  
FT CHAIN 1  
FT CHAIN 23  
FT CHAIN 24  
FT CHAIN 245  
FT CHAIN 246  
FT CHAIN 491  
FT CHAIN 492  
FT CHAIN 794  
FT CHAIN 795  
FT CHAIN 901  
FT CHAIN 900  
FT CHAIN 1087  
FT CHAIN 1088  
FT NON\_TER 1161  
SO SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 101; DB 12; Length 1161;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||

DB 828 FSOAKISLFYTEHEIMKFS 847

RESULT 5  
ID 09WMA2 PRELIMINARY; PRT; 2216 AA.

AC 09WMA2;  
DT 01-NOV-1999 (TEMBLrel. 12, Created)  
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AH3;



RX MEDLINE-21386014; PubMed-11495028;  
RA Fujiiwara K., Yokosuka O., Fukui K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020566; BAA35104.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
SQ SEQUENCE 2216 AA; 250209 MM; IA9D93FEC21FBE82 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2216;  
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTERHEIMKFS 20  
DB 834 FSOAKISLFYTERHEIMKFS 853  
|||||

RESULT 6  
067824 PRELIMINARY; PRT; 2218 AA.

ID 067824  
AC 067824;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GEM/RRHK RNA.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GBM;  
RX MEDLINE-94076453; PubMed-8254770;  
RA Graff J., Normann A., Feinstein S.M., Flemmig B.;  
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
to two cell culture adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75214; CA53024.1; -.  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.  
FT CHAIN 792 1422 P2 STRUCTURAL PROTEINS.  
FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.  
SQ SEQUENCE 2218 AA; 250502 MM; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTERHEIMKFS 20  
DB 828 FSOAKISLFYTERHEIMKFS 847  
|||||

RESULT 7  
067817 PRELIMINARY; PRT; 2218 AA.

ID 067817  
AC 067817;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE COMPLETE GENOME.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F.G.;  
RX MEDLINE-9381623; PubMed-7653108;  
RA Beneduce F., Pisan G., Divizia M., Pana A., Morace G.;  
RT "Complete nucleotide sequence of a cytopathic hepatitis A virus strain  
isolated in Italy.";  
RL Virus Res. 36:299-309(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F.G.;  
RA Morace G.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: X83302; CA58281.1; -.  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
SQ SEQUENCE 2218 AA; 250476 MM; 813B21D3E4E533CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTERHEIMKFS 20  
DB 828 FSOAKISLFYTERHEIMKFS 847  
|||||

RESULT 8  
09DL32 PRELIMINARY; PRT; 2225 AA.

ID 09DL32  
AC 09DL32;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L-A-1;  
RA Wang P.F., Jiang C.L., Liu J.Y., Zhang H.Y.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF314208; AAG45423.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
SQ SEQUENCE 2225 AA; 251297 MM; EBACE41B043E5E9B CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2225;  
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTERHEIMKFS 20

```
DB 834 FSOAKISLFTTEHEIMKFS 853
RESULT 9
ID 067825 PRELIMINARY; PRT; 2227 AA.
AC 067825;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GBM/WT RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GBM:
MEDLINE=94076453; PubMed=8254770;
RA Graft J., Norman A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
to two cell culture adapted variants.";
RL J. Virol. 68:548-554(1994).
EMBL: X75215; CAA53025.1;
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251563 MM; 4C4D79D352F936B4 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTTEHEIMKFS 20
DB 834 FSOAKISLFTTEHEIMKFS 853
RESULT 10
ID 067826 PRELIMINARY; PRT; 2227 AA.
AC 067826;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GBM/HFS RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GBM:
MEDLINE=94076453; PubMed=8254770;
RA Graft J., Norman A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
to two cell culture adapted variants.";
RL J. Virol. 68:548-554(1994).
EMBL: X75215; CAA53026.1;
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PROSITE: PS00910; RNA_helicase. 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
```

```
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251496 MM; 488CB7C962319457 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTTEHEIMKFS 20
DB 834 FSOAKISLFTTEHEIMKFS 853
RESULT 11
ID 09WMA4 PRELIMINARY; PRT; 2227 AA.
AC 09WMA4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH1:
MEDLINE=21386014; PubMed=11495028;
RA Fujiwara K., Yokosuka O., Fukui K., Imazeki F., Saisho H., Omata M.;
RT "Analysis of full-length hepatitis A virus genome in sera from
patients with fulminant and self-limited acute type A hepatitis.";
RL J. Hepatol. 35:112-119(2001).
EMBL: AB020564; BAA35102.1;
DR MEROPS: C03.005;
DR InterPro: IPR004004; Calic_pol_hel.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR Pfam: PF00910; RNA_helicase. 1.
DR PRINTS: PR00918; CALICVIRUSNS.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 2227 AA; 251304 MM; 0DEF6D2AEC29C0CE CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6, 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTTEHEIMKFS 20
DB 834 FSOAKISLFTTEHEIMKFS 853
RESULT 12
ID 09WMA3 PRELIMINARY; PRT; 2227 AA.
AC 09WMA3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH2:
MEDLINE=21386014; PubMed=11495028;
```

RA Fujiiwara K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
RT patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020565; BAA35103.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCC1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCC1\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251440 MW; E04F846CEC7F50FD CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTEHEIMKFS 20  
|||||  
DB 834 FSOAKISLFTEHEIMKFS 853

RESULT 13

ID O9WMA1 PRELIMINARY; PRT; 2227 AA.  
AC O9WMA1;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH1;  
RX MEDLINE-21386014; PubMed-11495028;  
RA Fujiiwara K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
RT patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020567; BAA35105.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCC1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCC1\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251415 MW; F92C8E2323FC5621 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTEHEIMKFS 20  
|||||  
DB 834 FSOAKISLFTEHEIMKFS 853

RESULT 14

ID O9WMAO PRELIMINARY; PRT; 2227 AA.  
AC O9WMAO;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH2;  
RX MEDLINE-21386014; PubMed-11495028;  
RA Fujiiwara K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
RT patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020568; BAA35106.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCC1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCC1\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251118 MW; 53E86B4432127E9B CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSOAKISLFTEHEIMKFS 20  
|||||  
DB 834 FSOAKISLFTEHEIMKFS 853

RESULT 15

ID O9WM99 PRELIMINARY; PRT; 2227 AA.  
AC O9WM99;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH3;  
RX MEDLINE-21386014; PubMed-11495028;  
RA Fujiiwara K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
RT patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020569; BAA35107.1; -.  
DR MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCC1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCC1\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251387 MW; 817640187672A23D CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||  
Db 834 FSOAKISLFYTEHEIMKFS 853

## RESULT 16

087092 PRELIMINARY; PRT; 184 AA.  
AC 087092;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE SEGMENT 2A-ENCODED PROTEIN (FRAGMENT).  
OS Simian hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
ON NCBI\_TaxID=12101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90012354; PubMed=2552172;  
RA Brown E.A., Jansen R.W., Lemon S.M.;  
RT "Characterization of a Simian hepatitis A virus (HAV): Antigenic and  
genetic comparison with human HAV.";  
RL J. Virol. 63:4932-4937(1989).  
DR EMBL: M34085; AAA47483.1; -.  
FT NON\_TER 1  
FT NON\_TER 184  
SQ SEQUENCE 184 AA; 21322 MW; 9FA851FA964A88A8 CRC64;

Query Match 97.0%; Score 98; DB 12; Length 184;  
Best Local Similarity 95.0%; Pred. No. 1.4e-08;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||  
Db 43 FSOAKISLFYTEHEIMKFS 62

## RESULT 17

091FH5 PRELIMINARY; PRT; 2227 AA.  
AC 091FH5;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
ON NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=HAF-203;  
RA Baptista M.L., Silva M., de Lima M.A., Yoshida C.F., Gaspar A.M.,  
Pires Lopes M.O., Galler R.;  
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated  
in Brazil and expression of the VP1 gene in a bacterial system.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF268396; AAF80114.1; -.  
FT MEROPS: C03.005; -.  
DR InterPro: IPR004004; Calic1\_pol\_hel.  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00918; CALICVIRUSNS.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251432 MW; 81913AEC8CA04200 CRC64;

Query Match 97.0%; Score 98; DB 12; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 2e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
|||||  
Db 834 FSOAKISLFYTEHEIMKFS 853

## RESULT 18

09ENR1 PRELIMINARY; PRT; 251 AA.  
AC 09ENR1;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
ON NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=A1;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB047652; BAB12160.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28749 MW; 58A520D873893445 CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 AKISLFYTEHEIMKFS 20  
|||||  
Db 1 AKISLFYTEHEIMKFS 17

## RESULT 19

09ENR0 PRELIMINARY; PRT; 251 AA.  
AC 09ENR0;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
ON NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=A10;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB047653; BAB12161.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28731 MW; B6C082013139BAF8 CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 AKISLFYTEHEIMKFS 20  
|||||  
Db 1 AKISLFYTEHEIMKFS 17

RESULT 20  
Q9EN09 PRELIMINARY; PRT; 251 AA.  
AC Q9EN09; 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A159;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047654; BAB12162.1; -  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28720 MW; C334248282F19CA CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
ID |||||||||||||||  
Db 1 AKISLFYTEHEIMKFS 17

RESULT 21  
Q9EN08 PRELIMINARY; PRT; 251 AA.  
AC Q9EN08; 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A160;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047655; BAB12163.1; -  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28693 MW; C32AD5651506751A CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
ID |||||||||||||||  
Db 1 AKISLFYTEHEIMKFS 17

RESULT 22  
Q9EN07 PRELIMINARY; PRT; 251 AA.  
AC Q9EN07; 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A161;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047656; BAB12164.1; -  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28766 MW; A0EC02E3609C4D5B CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
ID |||||||||||||||  
Db 1 AKISLFYTEHEIMKFS 17

RESULT 23  
Q9EN06 PRELIMINARY; PRT; 251 AA.  
AC Q9EN06; 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A162;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047657; BAB12165.1; -  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28699 MW; 56AA86DF60F6FD3B CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
ID |||||||||||||||  
Db 1 AKISLFYTEHEIMKFS 17

RESULT 24  
Q9EN05 PRELIMINARY; PRT; 251 AA.  
AC Q9EN05; 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A20;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047658; BAB1216.1; -.  
FT NON\_TER 1  
FT SEQUENCE 251 AA; 28720 MW; C34248282F19CA CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
|||  
Db 1 AKISLFYTEHEIMKFS 17

## RESULT 25

O9ENQ4 PRELIMINARY; PRT; 251 AA.  
ID O9ENQ4;  
AC O9ENQ4;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCB1\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A201;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047659; BAB12167.1; -.  
FT NON\_TER 1  
FT SEQUENCE 251 AA; 28720 MW; C34248282F19CA CRC64;

Query Match 85.1%; Score 86; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKISLFYTEHEIMKFS 20  
|||  
Db 1 AKISLFYTEHEIMKFS 17

Search completed: June 16, 2002, 00:08:53  
Job time: 791 sec



XX 18-APR-1997; 97WO-US06891.  
XX  
XX 19-APR-1996; 96US-0015644.  
XX  
XX (USSR ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fields HA, Khudyakov YE;  
XX  
XX WPI: 1997-535831/49.  
XX  
XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an  
XX immune response to HAV in a mammal or to detect the presence of  
XX antibodies against HAV in a mammal  
XX  
XX Claim 18; Page 112; 140pp; English.  
XX  
XX Peptides AAM42922-30 are immunogenic peptides corresponding to  
XX immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
XX substantially similar to a portion of the amino acid sequence of the P2A  
XX protein of HAV corresponding to amino acids 792-980. The present peptide  
XX is derived from amino acids 922-941, and has a reactivity of 31.3% with  
XX acute sera. Compositions containing the peptides can be used to induce an  
XX immune response to HAV in a mammal. The peptides can also be used to  
XX detect the presence of antibodies against HAV in mammalian serum. The  
XX peptides can also be used to make an antibody against HAV by  
XX administering the peptide to a mammal.  
XX  
XX Sequence 20 AA;  
XX  
XX

Query Match 100.0%; Score 104; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDEEIAANSKD 20  
1 kvnfpghmldleeiaanskd 20  
Db

RESULT 2  
AAB69444  
ID AAB69444 standard; Peptide; 21 AA.  
XX  
XX AAB69444;  
XX  
XX 20-APR-2001 (first entry)  
XX  
XX Synthetic HAV P2A peptide, SEQ ID NO: 44.  
XX  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XX antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
XX Hepatitis A virus.  
XX Synthetic.  
XX  
XX WO200105824-A2.  
XX  
XX 25-JAN-2001.  
XX  
XX 14-JUL-2000; 2000MO-US19267.  
XX  
XX 15-JUL-1999; 99US-0144412.  
XX  
XX (USSR ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fields HA, Khudyakov YE;  
XX  
XX WPI: 2001-112681/12.  
XX  
XX Synthetic peptides used as antigen sources for enzyme immunoassays  
XX detecting anti-hepatitis A virus and as vaccines -  
XX

PS Claim 13; Page 96; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
XX immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
XX comprise antigenic epitopes of the major structural capsid polypeptides  
XX or non-structural polypeptides of HAV with one or more glutamine  
XX molecules at the carboxy end of the peptide. The peptides are used to  
XX detect the presence of antibodies against HAV in mammalian serum, to  
XX detect the presence of HAV in a human or animal through the binding of  
XX the peptide to an antibody, to detect acute phase infection by detecting  
XX IgM antibodies in mammalian serum and detecting convalescence in a  
XX mammal. The peptides are used to detect or quantify HAV antibodies in  
XX samples in clinical or research-based assays using immunoblotting,  
XX fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
XX tracking of radioactive or bioluminescent markers, chromatography or  
XX electrophoresis. The peptides are used to induce an immune response to  
XX HAV when administered to a human or animal. Glutamine at the carboxy  
XX end of the peptides enhances the IgM antibody reactivity.  
XX  
XX Sequence 21 AA;  
XX  
XX

Query Match 100.0%; Score 104; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDEEIAANSKD 20  
1 kvnfpghmldleeiaanskd 20  
Db

RESULT 3  
AAP60066  
ID AAP60066 standard; Protein; 2227 AA.  
XX  
XX AAP60066;  
XX  
XX 26-JUN-1991 (first entry)  
XX  
XX Sequence of viral I434 polypeptide encoded by the complete  
XX nucleotide sequence of the HAV genome.  
XX  
XX Diagnosis; vaccine; passive immunotherapy.  
XX  
XX Hepatitis A virus.  
XX  
XX  
XX Key Location/Qualifiers  
XX Region 1..245  
XX Region /label= P1.1A  
XX Region 246..491  
XX Region /label= 1B  
XX Region 492..836  
XX Region /label= 1C  
XX Region 837..980  
XX Region /label= P2.2A  
XX Region 981..1076  
XX Region /label= 2B  
XX Region 1077..1422  
XX Region /label= 2C  
XX Region 1423..1484  
XX Region /label= P3.3A  
XX Region 1485..1507  
XX Region /label= 3B  
XX Region 1508..1678  
XX Region /label= 3C  
XX Region 1679..2227  
XX Region /label= 3D  
XX  
XX EPI99480-A.  
XX  
XX 29-OCT-1986.  
XX  
XX 03-APR-1986; 86EP-0302465.  
XX  
XX



[illegible]

```

PN  US4894228-A.
XX
XX  16-JAN-1990.
XX
XX  12-JUL-1988;      88US-0217824.
XX
XX  12-JUL-1988;      88US-0217824.
XX  12-JUL-1988;      88US-0652967.
XX
XX  (USSH ) US DEPT HEALTH & HUMAN.
XX
XX  Purell RH, Ticehurst JR, Cohen I, Emerson SU, Felstone SM;
PI  Daemer KJ, Gust ID:
XX
XX  WPI, 1990-075557/10.
XX  N-PSDB; AAQ03512.
XX
XX  Vaccine against hepatitis A virus infection - comprises novel
PT  attenuated hepatitis A virus strain.
XX
XX  Claim 1; Fig 1; 18pp; English.
XX
XX  The attenuated HAV is useful for inducing protective immunity against
CC  HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by
CC  several nucleotide changes distributed throughout the genome, is
CC  attenuated for chimpanzees, elicits serum neutralising antibodies, and is
CC  suitable for use as an HAV vaccine. It is noted that not all the changes
XX  are necessary for attenuation and use as a vaccine.
XX
XX  Sequence      2227 AA:

Query Match          100.0%; Score 104; DB 11; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1 KYNFPHGMLDEEIAANSKD 20
        |||||||
DB      922 kvnfhgmldleeianskd 941

RESULT  5
AAW34074 ID AAW34074 standard; Protein; 2227 AA.
XX
XX  AAW34074;
AC
XX  27-APR-1998 (first entry)
XX
XX  Hepatitis A virus HM-175 protein sequence.
DE
XX
XX  HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;
KM  infection; vaccine.
XX
XX  Hepatitis A virus HM-175.
OS
XX
XX  Key Location/Qualifiers
FH      1..23
FT      /label= VP4
FT      24..245
FT      /label= VP2
FT      246..491
FT      /label= VP3
FT      492..791
FT      /label= VP1
FT      792..980
FT      /label= 2A
FT      981..1087
FT      /label= 2B
FT      1088..1422
FT      /label= 2C
FT      1423..1496
FT      /label= 3A

```

FT	Protein	1497..1519	
FT		/label= 3B	
FT	Protein	1520..1738	
FT		/label= 3C	
FT	Protein	1739..2227	
FT		/label= 3D	
PN	W09740166-A2.		
XX			
PD	30-OCT-1997.		
XX			
PF	18-APR-1997;	97WO-US06506.	
XX			
PR	19-APR-1996;	96US-0015642.	
XX			
PA	(USSH ) US SEC DEPT HEALTH.		
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Emerson SU, Purcell RH, Raychaudhuri G;		
XX			
DR	WPI: 1997-535850/49.		
XX			
DR	N-PSDB: AAT93023.		
XX			
PT	Human attenuated HAV genome containing simian HAV 2C gene - useful		
PT	as vaccines against HAV infection		
XX			
PS	Disclosure: Flg 13A-D; 66pp; English.		
XX			
CC	This protein sequence is encoded by the human hepatitis A virus		
CC	(HAV) HM-175 wild-type genome (see AAT93023). Attenuated strain		
CC	HAV/7 is obtained by passage of HM-175 in African Green Monkey		
CC	kidney cells. A claimed DNA construct (1) comprises a genome of		
CC	HAV, where the genome is a human attenuated HAV genome in which a		
CC	region of the 2C gene has been replaced by a corresponding region		
CC	of a 2C gene of a simian AGM-27 HAV genome (see AAT93024). The		
CC	region of the 2C gene from AGM-27 contained in the construct		
CC	preferably encodes amino acids 120-328 of the 2C protein, amino		
CC	acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA		
CC	transcript of (1); (2) a cell transfected with (1) or the RNA		
CC	transcript of (1); (3) a HAV genome as above; (4) antibodies to the		
CC	HAV of (3); and (5) a host cell containing the HAV of (3). (1) or		
CC	its RNA transcript, can be used as a vaccine for preventing HAV in		
CC	a mammal. (1) or the RNA transcript can also be used to stimulate		
CC	the production of protective antibodies in the mammal.		
XX			
SQ	Sequence 2227 AA;		
XX			
XX	Query Match 100.0%; Score 104; DB 18; Length 2227;		
XX	Best Local Similarity 100.0%; Pred. No. 2e-08;		
XX	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 KVNPRGMLDEEIAANSKD 20		
XX			
DB	922 kvnfhgmldleeiaanskd 941		
XX			
XX	RESULT 6		
XX	AA018607		
XX	AA018607 standard; Protein: 2227 AA.		
XX	AA018607;		
XX			
DT	15-JAN-2001 (first entry)		
XX			
DE	Amino acid sequence of wild type Hepatitis A virus strain HM-175.		
XX	HAV, strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.		
XX	Hepatitis A virus.		
OS			
XX	US6113912-A.		
XX			
XX			

```

PD 05-SEP-2000.
PE 07-JUN-1995; 95US-0475886.
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93MO-US08610.
PR 10-MAR-1995; 95US-0397232.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
DR WPI: 2000-586464/55.
DR N-PSDB; AAA75476.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Fig 6A-K; 72pp; English.
XX
CC The present sequence is derived from a wild type hepatitis A virus
CC (HAV) strain HM-174. The sequence is modified to produce HAV which
CC are adapted to growth in the human fibroblast-like cell line MRC-5.
CC The HAV is able to propagate in MRC-5 cells and retain appropriate
CC attenuation. It is useful as a live vaccine for prophylaxis of
CC hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

Query Match 100.0%; Score 104; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMIDLEETANSKD 20
   ||||||||||||||||
DB 922 kvnphgmidleetaanskd 941

RESULT 7
AAB18608
ID AAB18608 standard; Protein: 2227 AA.
XX
AC AAB18608;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX
HW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW P-35 virus.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PE 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93MO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
DR WPI: 2000-586464/55.
DR N-PSDB; AAA75477.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

```

PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type -  
XX  
XX  
PS Disclosure; Columns 67-78; 72pp; English.  
XX  
CC The present sequence is derived from passage 35 of a wild type  
CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA;

Query Match 100.0%; Score 104; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNFPQMLDEEIAANSKD 20  
|||  
DB 922 kvnfpgmldleeiaanskd 941

RESULT 8  
AAB18609  
ID AAB18609 standard; Protein; 2227 AA.  
XX  
XX AAB18609;  
XX  
XX 15-JAN-2001 (first entry)  
XX  
XX Amino acid sequence of live attenuated Hepatitis A virus 4380.  
XX  
XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
XX HAV 4380.  
XX  
XX Hepatitis A virus.  
XX  
XX US6113912-A.  
XX  
XX 05-SEP-2000.  
XX  
XX 07-JUN-1995; 95US-0475886.  
XX  
XX 18-SEP-1992; 93US-0947338.  
XX 17-SEP-1993; 93MO-US08610.  
XX 10-MAR-1995; 95US-0397232.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AM;  
XX WPI: 2000-586464/55.  
XX DR N-PSDB; AAA75478.  
XX  
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
XX line useful as vaccine for protecting humans against hepatitis A virus  
XX infection, has modified genome compared to wild type -  
XX  
XX Disclosure; Columns 93-104; 72pp; English.  
XX  
XX The present sequence is derived from a live attenuated hepatitis A  
XX virus (HAV) of the invention, designated HAV 4380. The sequence is  
XX produced by modifying wild type HAV strain HM-174. The HAV of the  
XX invention are adapted to growth in the human fibroblast-like cell  
XX line MRC-5. The HAV is able to propagate in MRC-5 cells and retain  
XX appropriate attenuation. It is useful as a live vaccine for prophylaxis  
XX of hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA;

Query Match 100.0%; Score 104; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNFPQMLDEEIAANSKD 20  
|||  
DB 922 kvnfpgmldleeiaanskd 941

RESULT 9  
AAW42928  
ID AAW42928 standard; peptide; 20 AA.  
XX  
XX AAW42928;  
XX  
XX 28-APR-1998 (first entry)  
XX  
XX Immunogenic Hepatitis A virus peptide YK-1328.  
XX  
XX Immunogenic peptide; immunogenic epitope; p2A protein;  
XX immune response; antibody.  
XX  
XX Synthetic.  
XX OS Hepatitis A virus.  
XX PN WO9740147-A1.  
XX  
XX 30-OCT-1997.  
XX  
XX 18-APR-1997; 97MO-US06891.  
XX  
XX 19-APR-1996; 96US-0015644.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fields HA, Khudyakov YE;  
XX WPI: 1997-535831/49.  
XX  
XX Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an  
XX immune response to HAV in a mammal or to detect the presence of  
XX antibodies against HAV in a mammal  
XX  
XX Claim 18; Page 112; 140pp; English.  
XX  
XX Peptides AAW42922-30 are immunogenic peptides corresponding to  
XX immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
XX substantially similar to a portion of the amino acid sequence of the p2A  
XX protein of HAV corresponding to amino acids 792-980. The present peptide  
XX is derived from amino acids 931-950, and has a reactivity of 12.5% with  
XX acute sera. Compositions containing the peptides can be used to induce an  
XX immune response to HAV in a mammal. The peptides can also be used to  
XX detect the presence of antibodies against HAV in mammalian serum. The  
XX peptides can also be used to make an antibody against HAV by  
XX administering the peptide to a mammal.  
XX  
SQ Sequence 20 AA;

Query Match 51.0%; Score 53; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.043;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 DIEEIAANSKD 20  
|||  
DB 1 dieeiaanskd 11

RESULT 10  
AAB69445  
ID AAB69445 standard; Peptide; 21 AA.  
XX

AC AAB69445;  
XX  
DT 20-APR-2001 (first entry)  
XX  
XX Synthetic HAV P2A peptide, SEQ ID NO: 45.  
DE  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XX antigen; major structural capsid polypeptide; HAV antibody detection.  
KM  
XX Hepatitis A virus.  
OS  
XX Synthetic.  
XX  
XX WO200105824-A2.  
PN  
XX  
XX 25-JAN-2001.  
PD  
XX  
XX 14-JUL-2000; 2000WO-US19267.  
PF  
XX  
XX 15-JUL-1999; 99US-0144412.  
PR  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
XX  
XX Fields HA, Khudyakov YE.  
PI  
XX  
XX WPI; 2001-112681/12.  
DR  
XX  
XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines -  
PS  
XX  
XX Claim 13; Page 97; 130pp; English.  
CC The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
CC end of the peptides enhances the IgM antibody reactivity.  
CC  
XX  
SQ Sequence 21 AA;  
Query Match 51.0%; Score 53; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 DLEETIANSKD 20  
DB 1 dleetaanskd 11  
RESULT 11  
AAU32447  
ID AAU32447 standard; Protein: 195 AA.  
XX  
XX AAU32447;  
AC  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Novel human secreted protein #2938.  
DE  
XX  
XX Human; vaccination; gene therapy; nutritional supplement;  
KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX  
OS Homo sapiens.  
XX  
XX WO200179449-A2.  
PN  
XX  
XX 25-OCT-2001.  
PD  
XX  
XX 16-APR-2001; 2001WO-US08656.  
PF  
XX  
XX 18-APR-2000; 2000US-0552929.  
PR  
XX 26-JAN-2001; 2001US-0770160.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX  
XX WPI; 2001-611725/70.  
DR  
XX  
XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy -  
PT  
XX  
XX Claim 20; Page 615; 765pp; English.  
PS  
XX  
XX The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acids encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haematopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid  
CC sequences of novel human secreted proteins of the invention.  
CC  
XX  
SQ Sequence 195 AA;  
Query Match 43.3%; Score 45; DB 22; Length 195;  
Best Local Similarity 42.1%; Pred. No. 12;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
QY 2 VNFPHGMDLEETIANSKD 20  
DB 65 vspgpgvdlveevaaaspe 83  
RESULT 12  
AAG32341  
ID AAG32341 standard; Protein: 206 AA.  
XX  
XX AAG32341;  
AC  
XX  
XX 17-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 38996.  
DE  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX

PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR	28-SEP-1999;	9905-0156458.
PR	29-SEP-1999;	9905-0156596.
PR	04-OCT-1999;	9905-0157117.
PR	05-OCT-1999;	9905-0157753.
PR	06-OCT-1999;	9905-0157865.
PR	07-OCT-1999;	9905-0158029.
PR	08-OCT-1999;	9905-0158237.
PR	12-OCT-1999;	9905-0158369.
PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	13-OCT-1999;	9905-0159299.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161992.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

Query Match	43.3%	Score 45;	DB 21;	Length 206;
Best Local Similarity	44.4%	Freq. No. 13;		
Matches 8;	Conservative	5;	Mismatches	5;
			Indels	0;
			Gaps	0;
3 NFGHGLDLEETANSKD 20	::: :: ::			
32 dypvgllldlekllppgkd 49				
RESULT 13				
AAG10097				
AAG10097 standard; Protein; 242 AA.				
AC AAG10097;				
DT 17-OCT-2000 (first entry)				
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8283.				
XX Protein identification; signal transduction pathway; metabolic pathway; termination assay; genetic mapping; gene expression control; promoter;				
XX Arabidopsis thaliana.				
OS EPI033405-A2.				
PN				
XX 06-SEP-2000.				
PD				
PF 25-FEB-2000; 2000EP-0301439.				
XX				
PR 25-FEB-1999; 99US-0121825.				
PR 05-MAR-1999; 99US-0123180.				
PR 09-MAR-1999; 99US-0123548.				
PR 23-MAR-1999; 99US-0125788.				
PR 25-MAR-1999; 99US-0126264.				

PR	29-MAR-1999	9905-0126785
PR	01-APR-1999	9905-0127642
PR	06-APR-1999	9905-0128234
PR	08-APR-1999	9905-0128714
PR	16-APR-1999	9905-0129815
PR	19-APR-1999	9905-0130077
PR	21-APR-1999	9905-0130449
PR	22-APR-1999	9905-0130510
PR	28-APR-1999	9905-0131449
PR	30-APR-1999	9905-0132048
PR	30-APR-1999	9905-0132407
PR	04-MAY-1999	9905-0132484
PR	05-MAY-1999	9905-0132855
PR	06-MAY-1999	9905-0132867
PR	07-MAY-1999	9905-0132863
PR	11-MAY-1999	9905-0134256
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134741
PR	19-MAY-1999	9905-0134741
PR	21-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135553
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137702
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	18-JUN-1999	9905-0139452
PR	18-JUN-1999	9905-0139453
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139899
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142151
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-0142877
PR	13-JUL-1999	9905-0143242
PR	14-JUL-1999	9905-0143524
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144086

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145226.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156548.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 44.3%; Score 45; DB 21; Length 242;  
Best Local Similarity 44.4%; Pred. No. 15;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Oy 3 NPPHGMLEETIANSKD 20  
Db 68 dypvgldlekllppgkd 85  
::|:|:|:|:|

RESULT 14  
AAG32340  
ID AAG32340 standard; Protein: 242 AA.

AC AAG32340;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38995.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.  
PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142053.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.



PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 45; DB 21; Length 242;  
Best Local Similarity 44.4%; Pred. No. 15;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 3 NPPHGMLEIRANSKD 20  
:::|::|::|::|  
Db 68 dypvqldlekllppkd 85

RESULT 15  
AAG10096  
ID AAG10096 standard; Protein; 383 AA.  
XX AAG10096;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8282.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130831.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139869.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 43.3%; Score 45; DB 21; Length 383;  
Best Local Similarity 44.4%; Pred. No. 25;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
  
OY 3 NPHGMLDEETIAANSKD 20  
DB 209 dypvgldiekilppgkd 226  
  
RESULT 16  
AAG32339 standard; Protein; 383 AA.  
XX ID AAG32339;  
AC AAG32339;  
XX 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38994.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0133048.  
PR 30-APR-1999; 99US-01332407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135533.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140033.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144684.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155459.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.

PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.\*  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 43.3%; Score 45; DB 21; Length 383;  
Best Local Similarity 44.4%; Pred. No. 25;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
  
OY 3 NEPHCMLEETIANSKD 20  
Db 209 dypvgldlekllppkd 226  
  
RESULT 17  
AAG10095  
ID AAG10095 standard; Protein; 399 AA.  
XX AAG10095;  
XX AC  
XX AG10095;  
XX DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8281.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.

PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	05-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	06-AUG-1999;	9905-0147416
PR	09-AUG-1999;	9905-0147493
PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148361
PR	12-AUG-1999;	9905-0148535
PR	13-AUG-1999;	9905-0148684
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0149375
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0148426
PR	20-AUG-1999;	9905-0148722
PR	20-AUG-1999;	9905-0143723
PR	20-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149329
PR	25-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0153263
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154739
PR	22-SEP-1999;	9905-0155139
PR	23-SEP-1999;	9905-0155486
PR	06-OCT-1999;	9905-0155785
PR	07-OCT-1999;	9905-0157865
PR	08-OCT-1999;	9905-0158029
PR	12-OCT-1999;	9905-0158322
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0159293
PR	13-OCT-1999;	9905-0159294
PR	13-OCT-1999;	9905-0159295
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159330
PR	14-OCT-1999;	9905-0159531
PR	14-OCT-1999;	9905-0159637
PR	14-OCT-1999;	9905-0159638
PR	18-OCT-1999;	9905-0159584
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160989
PR	25-OCT-1999;	9905-0161404
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	26-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161920
PR	28-OCT-1999;	9905-0161992
PR	28-OCT-1999;	9905-0161993

```

PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 45; DB 21; Length 399;
Best Local Similarity 44.4%; Pred. No. 26;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0.

OY 3 NPPHGMLDLEETRANSKD 20
    . . . . .
DB 225 dypvgvlldekllppgkd 242
    . . . . .

```

RESULT	18
ABG11781	
ID	ABG11781 standard; Protein; 102 AA

AC ABG11781;

AC ABG11781;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #11772.

**KW** Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

**Homo sapiens.**

PN WO200175067-A2.

PD 11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX

XX

[illegible]

DR N-PSDB; AAS75968.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT responsible for genetic disorders or other traits and to assess

XX

XX

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC and gene mapping, and in recombinant production of (II). The

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC (II). (II) is useful for generating antibodies against it, detecting or

CC a food supplement. (II) and its blinding partners are useful in medical

CC disorders involving aberrant protein expression or biological activity.

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC diagnostic amino acid sequences of the invention.

CC specification, but was obtained in electronic format directly from WIPO

33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
8

sequence





PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149729.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158025.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.8%; Score 43.5; DB 21; Length 267;  
Best Local Similarity 45.0%; Pred. No. 30;  
Matches 9; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 KYNEPQML-DLEETANXK 19  
I: ||| |||: |  
Db 45 Kleinfmedledsgtgyk 64

RESULT 23  
AAG37736



ID AAG37736 standard; Protein; 267 AA.  
AC AAG37736;  
XX  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46449.  
XX  
KM Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
PN EP103405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 200EP-0301439.  
PF  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132485.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132683.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147319.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161923.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.8%; Score 43.5; DB 21; Length 267;  
Best Local Similarity 45.0%; Pred. No. 30;  
Matches 9; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 KYNPPHGM-DIEETANSK 19  
|:|:|:|:|:|:|:|:|:|  
Db 45 Kiefpmediedisgltgk 64

RESULT 24  
AAG24210  
ID AAG24210 standard; Protein; 269 AA.  
XX  
AC AAG24210;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment; SEQ ID NO: 27795.

XX XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142277.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.8%; Score 43.5; DB 21; Length 269;  
Best Local Similarity 45.0%; Pred. No. 30;  
Matches 9; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

OY 1 KVNPPHML-DLERIANSK 19  
|:|:| |:|:|  
Db 45 KieIphnefedledlsytk 64

RESULT 25

ID AAG37720 standard; Protein: 426 AA.

AC AAG37720;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46428.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 21-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.

PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142820.  
PR 12-JUL-1999; 990S-0142877.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148365.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151438.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.

PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158359.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 41.88; Score 43.5; DB 21; Length 426;  
 Best Local Similarity 45.0%; Pred. No. 51;  
 Matches 9; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
 OY 1 KVFPHGWL-DLEETIAANSK 19  
 | : | | | | | : | | : | : |  
 Db 204 Kieipnefedledalsgtgk 223

Search completed: June 16, 2002, 00:01:47  
 Job time: 12736 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:17 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-44  
Perfect score: 104  
Sequence: 1 KVNPHGMLEIFRANSKD 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :  
1: Issued Patents\_AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	104	100.0	2227	3	US-08-475-886-2	Sequence 2, Appl1
2	104	100.0	2227	3	US-08-475-886-4	Sequence 4, Appl1
3	104	100.0	2227	3	US-08-475-886-6	Sequence 6, Appl1
4	104	100.0	2227	4	US-08-397-232-2	Sequence 2, Appl1
5	104	100.0	2227	4	US-08-397-232-4	Sequence 4, Appl1
6	104	100.0	2227	2	US-09-171-387-2	Sequence 2, Appl1
7	45	43.3	417	2	US-09-099-677A-6	Sequence 6, Appl1
8	45	43.3	417	3	US-09-261-471-6	Sequence 6, Appl1
9	44	42.3	432	2	US-09-099-677A-3	Sequence 3, Appl1
10	44	42.3	432	3	US-09-261-471-3	Sequence 3, Appl1
11	43	41.3	98	4	US-09-181-487-2	Sequence 219, Appl
12	43	41.3	99	4	US-09-227-357-219	Sequence 31, Appl
13	39	37.5	579	1	US-08-126-564A-31	Sequence 31, Appl
14	39	37.5	579	5	PCT-US94-09143-31	Sequence 24, Appl
15	39	37.5	1066	4	US-09-004-838-24	Sequence 90, Appl
16	39	37.5	1333	4	US-09-004-838-90	Sequence 95, Appl
17	39	37.5	1604	4	US-09-004-838-95	Sequence 92, Appl
18	39	37.5	1805	4	US-09-004-838-92	Sequence 125, App
19	39	37.5	1817	4	US-09-004-838-125	Sequence 9, Appl1
20	38	36.5	1250	3	US-08-938-291A-9	Sequence 2, Appl1
21	38	36.5	1367	2	US-08-249-687C-2	Sequence 2, Appl1
22	38	36.5	1367	2	US-08-625-819-2	Sequence 2, Appl1
23	38	36.5	1367	3	US-08-746-559A-2	Sequence 18, Appl
24	38	36.5	1367	4	US-08-864-641B-18	Sequence 2, Appl1
25	37.5	35.6	568	4	US-09-238-373-2	Sequence 1, Appl1
26	37	35.6	823	1	US-08-261-465-1	Sequence 5, Appl1
27	37	35.6	823	1	US-08-405-254-5	

28	37	35.6	1525	4	US-09-396-651B-1	Sequence 1, Appl1
29	36.5	35.1	350	2	US-08-828-922-1	Sequence 1, Appl1
30	36.5	35.1	4544	2	US-08-469-486-52	Sequence 52, Appl
31	36.5	35.1	4544	2	US-08-469-658-52	Sequence 52, Appl
32	36	34.6	95	3	US-08-946-329A-86	Sequence 86, Appl
33	36	34.6	119	4	US-08-858-207A-394	Sequence 394, App
34	36	34.6	444	2	US-08-823-516-145	Sequence 145, App
35	36	34.6	479	4	US-09-446-504-64	Sequence 64, Appl
36	36	34.6	498	4	US-08-712-266-64	Sequence 64, Appl
37	36	34.6	498	4	US-08-511-485-13	Sequence 13, Appl
38	36	34.6	793	2	US-08-468-558-5	Sequence 5, Appl1
39	36	34.6	793	4	US-08-676-444-5	Sequence 5, Appl1
40	36	34.6	2004	1	US-08-375-709-15	Sequence 15, Appl
41	36	34.6	2004	1	US-08-752-929-15	Sequence 15, Appl
42	36	34.6	2004	4	US-09-090-793-9	Sequence 9, Appl1
43	35	33.7	246	2	US-08-825-782-1	Sequence 1, Appl1
44	35	33.7	246	1	US-08-286-888B-1	Sequence 1, Appl1
45	35	33.7	246	1	US-08-294-386C-1	Sequence 1, Appl1
46	35	33.7	246	1	US-08-294-386C-3	Sequence 3, Appl1
47	35	33.7	246	1	US-08-299-249A-1	Sequence 1, Appl1
48	35	33.7	246	1	US-08-299-249A-10	Sequence 10, Appl
49	35	33.7	246	5	PCT-US94-06543-1	Sequence 1, Appl1
50	35	33.7	246	5	PCT-US95-10224-3	Sequence 3, Appl1
51	35	33.7	246	5	PCT-US95-10224-3	Sequence 3, Appl1
52	35	33.7	292	1	US-08-286-888B-6	Sequence 6, Appl1
53	35	33.7	347	1	US-07-637-870-5	Sequence 5, Appl1
54	35	33.7	347	1	US-07-637-870-5	Sequence 10, Appl
55	35	33.7	348	1	US-07-637-399-8	Sequence 8, Appl1
56	35	33.7	348	1	US-08-112-703-8	Sequence 8, Appl1
57	35	33.7	374	1	US-08-625-876-2	Sequence 2, Appl1
58	35	33.7	374	1	US-08-625-876-4	Sequence 4, Appl1
59	35	33.7	374	1	US-08-625-876-6	Sequence 6, Appl1
60	35	33.7	374	1	US-08-625-876-6	

## ALIGNMENTS

RESULT 1  
US-08-475-886-2  
Sequence 2, Application US/08475886A  
Patent No. 6113912  
GENERAL INFORMATION:  
APPLICANT: FONKHOUER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US2  
CURRENT APPLICATION NUMBER: US/08/475,886A  
CURRENT FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 07/947,338  
EARLIER FILING DATE: 1992-09-18  
EARLIER APPLICATION NUMBER: 08/397,232  
EARLIER FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 104; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPHGMLEIFRANSKD 20  
|||||  
Db 922 KVNPHGMLEIFRANSKD 941

```
RESULT 2
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4
```

```
Query Match          100.0%; Score 104; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KVNPPHGMDLLEETIANSKD 20
|||||
```

```
DB 922 KVNPPHGMDLLEETIANSKD 941
```

```
RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6
```

```
Query Match          100.0%; Score 104; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KVNPPHGMDLLEETIANSKD 20
|||||
```

```
DB 922 KVNPPHGMDLLEETIANSKD 941
```

```
RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
```

```
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2
```

```
Query Match          100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KVNPPHGMDLLEETIANSKD 20
|||||
```

```
DB 922 KVNPPHGMDLLEETIANSKD 941
```

```
RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4
```

```
Query Match          100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KVNPPHGMDLLEETIANSKD 20
|||||
```

```
DB 922 KVNPPHGMDLLEETIANSKD 941
```

```
RESULT 6
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHAUDHURI, GOPA;
```



```
EMERSON, SUZANNE, U.;
PURCELL, ROBERT, H.
TITLE OF INVENTION: SIMIAN-HUMAN HAV
HAVING A CHIMERIC 2C PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/171,387
FILING DATE: 24-Mar-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCF/US97/06506
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US60/015,642
FILING DATE: 19-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feller
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4229U51
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 2227 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-09-171-387-2

Query Match      100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPHGMLDEETIAANSKD 20
DB 922 KVNPHGMLDEETIAANSKD 941

RESULT 7
US-09-099-677A-6
Sequence 6, Application US/09099677A
Patent No. 5965369
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: HUMAN SUCCINYL-COENZYME A SYNTHETASE HOLOENZYME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,677A
FILING DATE: June 18, 1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CERONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0545 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: g164669
US-09-099-677A-6

Query Match      43.3%; Score 45; DB 2; Length 417;
Best Local Similarity 42.1%; Pred. No. 7.8;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNPFGMLDEETIAANSKD 20
DB 153 VGSPOGGVDEEVAANSPE 171

RESULT 8
US-09-261-471-6
Sequence 6, Application US/09261471
Patent No. 6025123
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN SUCCINYL-COA SYNTHETASE HOLOENZYME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/261,471
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/099,677
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CERONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0545 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
```

STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: g164669  
US-09-261-471-6

Query Match 43.3%; Score 45; DB 3; Length 417;  
Best Local Similarity 42.1%; Pred. No. 7.8;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDEEIAANSKD 20  
DB 153 VGPOGGVDIEEVAASNP 171

## RESULT 9

US-09-099-677A-3  
Sequence 3, Application US/09099677A  
Patent No. 5965369  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neel C.  
TITLE OF INVENTION: HUMAN SUCCINYL-COENZYME A SYNTHETASE HOLOENZYME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,677A  
FILING DATE: June 18, 1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0545 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 432 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSBPT06  
CLONE: 3273853  
US-09-099-677A-3

Query Match 42.3%; Score 44; DB 2; Length 432;  
Best Local Similarity 42.1%; Pred. No. 12;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDEEIAANSKD 20  
DB 168 VGPOGGVDIEEVAASNP 186

## RESULT 10

US-09-261-471-3  
Sequence 3, Application US/09261471  
Patent No. 6025123  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neel C.  
TITLE OF INVENTION: HUMAN SUCCINYL-COA SYNTHETASE HOLOENZYME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/261,471  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/099,677  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0545 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 432 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSBPT06  
CLONE: 3273853  
US-09-261-471-3

Query Match 42.3%; Score 44; DB 3; Length 432;  
Best Local Similarity 42.1%; Pred. No. 12;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDEEIAANSKD 20  
DB 168 VGPOGGVDIEEVAASNP 186

## RESULT 11

US-09-181-487-2  
Sequence 2, Application US/09181487  
Patent No. 6165752  
GENERAL INFORMATION:  
APPLICANT: DUCKWORTH, DAVID MALCOLM  
TITLE OF INVENTION: NOVEL COMPOUNDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ratner & Prestia  
STREET: P.O. Box 980  
CITY: Valley Forge  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/181,487
FILING DATE: 28-OCT-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9807082.4
FILING DATE: 01-APR-1998
APPLICATION NUMBER: 9815489.1
FILING DATE: 16-JUL-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30094
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 98 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-181-487-2

Query Match      41.3%  Score 43;  DB 4;  Length 98;
Best Local Similarity 44.4%  Pred. No. 3.2;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      3  NPPGMDLEIEIANSKD 20
      1 1 1 1 1 1 1 1 1 1
Db      40  NLPGEVADLTQIDVNVOD 57

RESULT 12
US-09-227-357-219
Sequence 219, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357
EARLIER FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08

```

```

EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 219
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (99)
OTHER INFORMATION: Xaa equals stop translation
US-09-227-357-219

Query Match      41.3%  Score 43;  DB 4;  Length 99;
Best Local Similarity 44.4%  Pred. No. 3.2;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      3  NPPGMDLEIEIANSKD 20
      1 1 1 1 1 1 1 1 1 1
Db      40  NLPGEVADLTQIDVNVOD 57

RESULT 13
US-08-126-564A-31
Sequence 31, Application US/08126564A
Patent No. 5436150
GENERAL INFORMATION:
APPLICANT: Chandrasegaran, Srinivasan
TITLE OF INVENTION: Functional Domains in POKI
TITLE OF INVENTION: Restriction Endonuclease
NUMBER OF SEQUENCES: 48

```

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Cushman, Darby & Cushman  
;; STREET: 1100 New York Ave., N.W.  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-3918  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0,  
;; SOFTWARE: Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/126,564A  
;; FILING DATE: 27-SEPTEMBER-93  
;; CLASSIFICATION: 435  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kokulis, Paul N.  
;; REGISTRATION NUMBER: 16,773  
;; REFERENCE/DOCKET NUMBER: PNK/4130/82506/CLB  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-861-3503  
;; TELEFAX: 202-822-0944  
;; TELEX: 6714627 CUSH  
;;  
;; INFORMATION FOR SEQ ID NO: 31:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 579 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-126-564A-31

Query Match 37.5% Score 39; DB 1; Length 579;  
Best Local Similarity 57.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 5 PHGMJLDEIRANS 18  
||:::|||11  
Db 405 PHEYTELEIRANS 418

RESULT 14  
PCT-US94-09143-31  
; Sequence 31, Application PC/TUS9409143  
; GENERAL INFORMATION:  
; APPLICANT: Chandrasegaran, Srinivasan  
; TITLE OF INVENTION: Functional Domains in Foki  
; TITLE OF INVENTION: Restriction Endonuclease  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Cushman, Darby & Cushman  
; STREET: 1100 New York Ave., N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0,  
; SOFTWARE: Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/09143  
; FILING DATE: 23-AUG-1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/126,564  
; FILING DATE: 27-SEPTEMBER-93  
; ATTORNEY/AGENT INFORMATION:

;; NAME: Kokulis, Paul N.  
;; REGISTRATION NUMBER: 16,773  
;; REFERENCE/DOCKET NUMBER: PNK/4130/82506/CLB  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-861-3503  
;; TELEFAX: 202-822-0944  
;; TELEX: 6714627 CUSH  
;;  
;; INFORMATION FOR SEQ ID NO: 31:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 579 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; PCT-US94-09143-31

Query Match 37.5% Score 39; DB 5; Length 579;  
Best Local Similarity 57.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 5 PHGMJLDEIRANS 18  
||:::|||11  
Db 405 PHEYTELEIRANS 418

RESULT 15  
US-09-004-838-24  
; Sequence 24, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:  
; APPLICANT: Michelmore, Richard W.  
; APPLICANT: Meyers, Blake  
; TITLE OF INVENTION: Procedures and Materials for  
; TITLE OF INVENTION: Confering Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 140  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/004,838  
; FILING DATE: 09-JAN-1998  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/781,734  
; FILING DATE: 10-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Einhorn, Gregory P.  
; REGISTRATION NUMBER: 38,440  
; REFERENCE/DOCKET NUMBER: 023070-078810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0300  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1066 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..1066

OTHER INFORMATION: /note= "R1G2B amino acids"  
US-09-004-838-24

Query Match 37.5%: Score 39; DB 4; Length 1066;  
Best Local Similarity 53.8%: Pred. No. 2.4e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HGMNLEDEVIANS 18  
||| |||::: |  
DB 752 HGMNLEDEVEVKS 764

RESULT 16  
US-09-004-838-90

; Sequence 90, Application US/09004838  
; Patent No. 6350933

; GENERAL INFORMATION:

; APPLICANT: Michelmore, Richard W.

; APPLICANT: Shen, Kathy

; TITLE OF INVENTION: Procedures and Materials for

; NUMBER OF SEQUENCES: 140

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/004,838

; FILING DATE: 09-JAN-1998

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/781,734

; FILING DATE: 10-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Einhorn, Gregory P.

; REGISTRATION NUMBER: 38,440

; REFERENCE/DOCKET NUMBER: 023070-078810US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 90:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1323 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: -

; LOCATION: 1..1323

; OTHER INFORMATION: /note= "RG2B deduced sequence"

US-09-004-838-90

Query Match 37.5%: Score 39; DB 4; Length 1323;

Best Local Similarity 53.8%: Pred. No. 3e+02;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HGMNLEDEVIANS 18  
||| |||::: |  
DB 752 HGMNLEDEVEVKS 764

RESULT 17  
US-09-004-838-95

; Sequence 95, Application US/09004838

; Patent No. 6350933

; GENERAL INFORMATION:

; APPLICANT: Michelmore, Richard W.

; APPLICANT: Shen, Kathy

; TITLE OF INVENTION: Procedures and Materials for

; NUMBER OF SEQUENCES: 140

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/004,838

; FILING DATE: 09-JAN-1998

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/781,734

; FILING DATE: 10-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Einhorn, Gregory P.

; REGISTRATION NUMBER: 38,440

; REFERENCE/DOCKET NUMBER: 023070-078810US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 95:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1604 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: -

; LOCATION: 1..1604

; OTHER INFORMATION: /note= "RG2D deduced sequence"

US-09-004-838-95

Query Match 37.5%: Score 39; DB 4; Length 1604;

Best Local Similarity 53.8%: Pred. No. 3.8e+02;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HGMNLEDEVIANS 18  
||| |||::: |  
DB 750 HGMNLEDEVEVKS 762

RESULT 18  
US-09-004-838-92

; Sequence 92, Application US/09004838

; Patent No. 6350933

; GENERAL INFORMATION:

; APPLICANT: Michelmore, Richard W.

; APPLICANT: Shen, Kathy

; TITLE OF INVENTION: Procedures and Materials for

; NUMBER OF SEQUENCES: 140

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,838  
FILING DATE: 09-JAN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/781,734  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 023070-078810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 92:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1805 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..1805  
OTHER INFORMATION: /note= "RG2C deduced sequence"  
US-09-004-838-92

Query Match 37.5%; Score 39; DB 4; Length 1805;  
Best Local Similarity 53.8%; Pred. No. 4.4e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 GNMIDLEFIANS 18  
||| |||: |  
DB 750 GNMIDLEVEVKS 762

RESULT 19  
US-09-004-838-125  
Sequence 125, Application US/09004838  
Patent No. 6350933  
GENERAL INFORMATION:  
APPLICANT: Michelmore, Richard W.  
APPLICANT: Shen, Kathy  
TITLE OF INVENTION: Procedures and Materials for  
NUMBER OF INVENTION: Conferring Pest Resistance in Plants  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,838  
FILING DATE: 09-JAN-1998

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/781,734  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 023070-078810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 125:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1817 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..1817  
OTHER INFORMATION: /note= "RG2S deduced sequence"  
US-09-004-838-125

Query Match 37.5%; Score 39; DB 4; Length 1817;  
Best Local Similarity 53.8%; Pred. No. 4.4e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 GNMIDLEFIANS 18  
||| |||: |  
DB 750 GNMIDLEVEVKS 762

RESULT 20  
US-08-938-291A-9  
Sequence 9, Application US/08938291A  
Patent No. 6117673  
GENERAL INFORMATION:  
APPLICANT: Lev, Sima  
APPLICANT: Plowman, Gregory D.  
TITLE OF INVENTION: RDBG PROTEINS AND RELATED  
NUMBER OF INVENTION: PRODUCTS AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,291A  
FILING DATE: September 26, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,337  
FILING DATE: October 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 228/172  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440

TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1250 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-938-291A-9

Query Match 36.5%; Score 38; DB 3; Length 1250;  
Best Local Similarity 45.5%; Pred. No. 4,7e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 NPPGMLDLEE 13  
|||||:  
DB 1120 NPPGMLISPAD 1130

RESULT 21  
US-08-249-687C-2  
Sequence 2, Application US/08249687C  
Patent No. 5942412  
GENERAL INFORMATION:  
APPLICANT: PRAGER, DIANE  
APPLICANT: MELMED, SHLOMO  
TITLE OF INVENTION: POLYNUCLEIC ACID ENCODING  
TITLE OF INVENTION: VARIANT INSULIN-LIKE GROWTH FACTOR I RECEPTOR BETA  
TITLE OF INVENTION: SUBUNIT & RECEPTOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder & Poplawski  
STREET: 444 South Flower St., 19th Floor  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/249,687C  
FILING DATE: 26-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/044,540  
FILING DATE: 06-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Basille, Lena  
REGISTRATION NUMBER: P-44,026  
REFERENCE/DOCKET NUMBER: P07 32349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213-622-7700  
TELEFAX: 213-489-4210  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1367 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-249-687C-2

Query Match 36.5%; Score 38; DB 2; Length 1367;  
Best Local Similarity 37.5%; Pred. No. 4,7e+02;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
OY 4 FPHGMLDEETANSK 19

DB 673 YADGTIDIEVTENPK 688  
:|:|:|:|

RESULT 22  
US-08-625-819-2  
Sequence 2, Application US/08625819  
Patent No. 5958872  
GENERAL INFORMATION:  
APPLICANT: O'CONNOR, Rosemary; and  
APPLICANT: BASERGA, Renato L.  
TITLE OF INVENTION: ACTIVE SURVIVAL DOMAINS OF IGF-IR  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE and DORR LLP  
STREET: 1455 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,819  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WIXON, Henry N.  
REGISTRATION NUMBER: 32,073  
REFERENCE/DOCKET NUMBER: 104322.162  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 942-8459  
TELEFAX: (202) 942-8484  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1367 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-625-819-2

Query Match 36.5%; Score 38; DB 2; Length 1367;  
Best Local Similarity 37.5%; Pred. No. 4,7e+02;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 4 FPHGMLDEETANSK 19  
:|:|:|:|  
DB 673 YADGTIDIEVTENPK 688

RESULT 23  
US-08-746-559A-2  
Sequence 2, Application US/08746559A  
Patent No. 6084085  
GENERAL INFORMATION:  
APPLICANT: Renato Baserga  
APPLICANT: Mariana Resnikoff  
APPLICANT: Consuelo D'Ambrosio  
APPLICANT: Andre Ferber  
TITLE OF INVENTION: Method of Inducing Resistance to Tumor Growth  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6084085rls LLP  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103

```
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,559A
FILING DATE: 13-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/006,699
FILING DATE: 14-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Legaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: TJU-2063
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1367 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-746-559A-2
```

```
Query Match          36.5%; Score 38; DB 3; Length 1367;
Best Local Similarity 37.5%; Pred. No. 4.7e+02;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
```

```
QY      4  FPHGMDLEIEIANSK 19
DB      673 YADGTIDIEEVENPK 688
```

```
RESULT 24
US-08-864-641B-18
; Sequence 18, Application US/08864641B
; Patent No. 6312684
; GENERAL INFORMATION:
; APPLICANT: Baserga, Renato
; APPLICANT: Abraham, David
; APPLICANT: Resnicoff, Mariana
; TITLE OF INVENTION: Method Of Inducing Resistance To Tumor Growth
; FILE REFERENCE: TJU2137
; CURRENT APPLICATION NUMBER: US/08/864,641B
; CURRENT FILING DATE: 1997-05-29
; PRIOR APPLICATION NUMBER: 08/340,732
; PRIOR FILING DATE: 1994-11-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: No. 6312684el Sequence
US-08-864-641B-18
```

```
Query Match          36.5%; Score 38; DB 4; Length 1367;
Best Local Similarity 37.5%; Pred. No. 4.7e+02;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
```

```
QY      4  FPHGMDLEIEIANSK 19
DB      673 YADGTIDIEEVENPK 688
```

```
RESULT 25
US-09-238-373-2
```

```
; Sequence 2, Application US/09238373A
; Patent No. 6187562
; GENERAL INFORMATION:
; APPLICANT: DICKMORTH, DAVID MALCOLM
; APPLICANT: GODDEN, ROBERT JAMES
; APPLICANT: TESTA, TANIA TARSON
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30034
; CURRENT APPLICATION NUMBER: US/09/238,373A
; CURRENT FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: UK 9824026.0
; EARLIER FILING DATE: 1998-11-03
; EARLIER APPLICATION NUMBER: EP 98300625.5
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-238-373-2
```

```
Query Match          36.1%; Score 37.5; DB 4; Length 568;
Best Local Similarity 53.8%; Pred. No. 2.1e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
```

```
QY      4  FPHGMD-LEEIA 15
DB      290 FPHGVDPVEVA 302
```

```
Search completed: June 16, 2002, 00:03:19
Job time: 8533 sec
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

# OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:20 ; Search time 108.75 Seconds  
(without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-432a-44

Perfect score: 104  
Sequence: 1 KVNPFHGMLEETIAANSKD 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	2227	1 GNNYHM	genome polypotein
2	104	100.0	2227	1 GNNYHM	genome polypotein
3	104	100.0	2227	1 GNNYHM	genome polypotein
4	104	100.0	2227	1 GNNYHM	genome polypotein
5	101	97.1	2230	1 GNNYHM	genome polypotein
6	47	45.2	333	2 A12485	hypothetical prote
7	47	45.2	387	2 B86669	RecA protein (lipo
8	47	45.2	506	1 S58522	glycine--tRNA 11ga
9	47	45.2	1060	2 F88710	protein C015.4 [1
10	47	45.2	1079	2 T30996	hypothetical prote
11	46	44.2	259	2 F82087	deoxyribose-phosph
12	46	44.2	444	2 C89768	conserved hypotet
13	46	44.2	677	2 H64574	DNA topoisomerase
14	45	43.3	195	2 T08812	probable succinate
15	45	43.3	383	2 F96582	hypothetical prote
16	45	43.3	417	2 A44529	succinate--CoA lig
17	45	43.3	852	2 B72685	hypothetical prote
18	45	43.3	935	2 T19011	hypothetical prote
19	45	43.3	1213	2 T19835	hypothetical prote
20	44	42.3	223	2 T37962	hypothetical prote
21	44	42.3	344	2 D97761	hypothetical prote
22	44	42.3	500	2 S50508	AMP1 protein - yea
23	43.5	41.8	373	2 A69773	hypothetical prote
24	43	41.3	140	2 E81659	conserved hypotet
25	43	41.3	141	2 G71501	hypothetical prote
26	43	41.3	257	2 AF2592	2-deoxyribose-5-ph
27	43	41.3	259	2 H97374	2-deoxyribose-5-ph
28	43	41.3	309	2 H71089	hypothetical prote
29	43	41.3	327	2 T00876	probable glycerate

30	43	41.3	440	2 B71858	adenylosuccinate 1
31	43	41.3	440	2 H64658	adenylosuccinate 1
32	43	41.3	609	2 AB0955	glutamine--fructos
33	43	41.3	917	2 T21870	hypothetical prote
34	42.5	40.9	599	2 A86810	1-deoxyxylulose-5-
35	42	40.4	222	2 C82343	conserved hypotet
36	42	40.4	229	2 S77449	hydrogenase expres
37	42	40.4	259	2 D91296	2-deoxyribose-5-ph
38	42	40.4	259	2 F86137	2-deoxyribose-5-ph
39	42	40.4	290	2 S19426	hypothetical prote
40	42	40.4	309	2 T33259	hypothetical prote
41	42	40.4	342	2 C72313	hypothetical prote
42	42	40.4	365	2 B48945	hypothetical prote
43	42	40.4	384	2 E82088	recombination prot
44	42	40.4	495	2 D64578	conserved hypotet
45	42	40.4	539	2 G70520	probable csp prote
46	42	40.4	839	2 C97250	preprotein translo
47	42	40.4	1742	2 S24600	projectin - fruit
48	42	40.4	6658	2 G64187	anthranilate synth
49	41	39.4	193	2 T13931	hypothetical prote
50	41	39.4	218	2 D71693	hypothetical prote
51	41	39.4	274	2 T36347	hypothetical prote
52	41	39.4	365	2 T20652	hypothetical prote
53	41	39.4	400	2 C34443	nitrogenase cofact
54	41	39.4	400	2 AE1988	nitrogenase cofact
55	41	39.4	607	2 B84153	two-component sens
56	41	39.4	834	2 T19010	hypothetical prote
57	41	39.4	848	2 G86708	aconitate hydratase
58	40.5	38.9	237	1 B8AC58	virB8 protein Ag
59	40.5	38.9	237	2 AD3249	component of type
60	40.5	38.9	434	2 E71638	UDP-glucose 6-dehy

## ALIGNMENTS

RESULT 1  
GNNYHM  
genome polypotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polypotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 104; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNPFHGMLEETIAANSKD 20

```
Db      922  KVNFPHGMDLEBIANSKD  941
|||||
RESULT  2
GNMYHR
genome polypeptide - human hepatitis A virus
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro
NA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03903
R:Naftarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985
A:Title: Primary structure and gene organization of human hepatitis A virus.
A:Reference number: A03903; MUID:85190549
A:Accession: A03903
A:Molecule type: genomic RNA
A:Residues: 1-2227 <MAN>
A:Cross-references: GB:K02990; NID:g329596; PIDN:AAA5472.1; PID:g329597
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans
F:1-245/Product: coat protein 1A #status predicted <C1A>
F:1246-491/Product: coat protein 1B #status predicted <C1B>
F:492-836/Product: coat protein 1C #status predicted <C1C>
F:837-980/Product: core protein 2A #status predicted <C2A>
F:981-1076/Product: core protein 2B #status predicted <C2B>
F:1077-1422/Product: core protein 2C #status predicted <C2C>
F:1423-1484/Product: protein 3A #status predicted <C3A>
F:1485-1507/Product: protein 3B #status predicted <C3B>
F:1508-1578/Product: cysteine proteinase, protein 3C #status predicted <C3C>
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match      100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  KVNFPHGMDLEBIANSKD  20
|||||
Db      922  KVNFPHGMDLEBIANSKD  941

RESULT  3
GNMYHR
genome polypeptide - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro
NA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A94149; A25914; A94508
R:Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with
A:Reference number: A94149; MUID:8715701
A:Accession: A94149
A:Status: nucleic acid sequence not shown
A:Molecule type: genomic RNA
A:Residues: 1-2227 <COH>
A:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA5471.1; PID:g329595
A:Note: submitted to Genbank, August 1987
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans
F:1-245/Product: coat protein 1A #status predicted <F1A>
F:1246-491/Product: coat protein 1B #status predicted <F1B>
F:492-836/Product: coat protein 1C #status predicted <F1C>
F:837-980/Product: core protein 2A #status predicted <F2A>
F:981-1076/Product: core protein 2B #status predicted <F2B>
F:1077-1422/Product: core protein 2C #status predicted <F2C>
F:1423-1484/Product: protein 3A #status predicted <F3A>
F:1485-1507/Product: protein 3B #status predicted <F3B>
```

```
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <F3C>
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <F3D>

Query Match      100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  KVNFPHGMDLEBIANSKD  20
|||||
Db      922  KVNFPHGMDLEBIANSKD  941

RESULT  4
GNMYHR
genome polypeptide - human hepatitis A virus (strain MBB)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
VpG; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-May-1996
C:Accession: J50303
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wisel, T.; Klein, R.; Wimmer, E.; Deinhard
Virus Res. 8, 153-171, 1987
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso
A:Reference number: J50303; MUID:88045071
A:Accession: J50303
A:Molecule type: genomic RNA
A:Residues: 1-2227 <PAN>
A:Cross-references: EMBL:M20273
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h
F:1-245/Product: coat protein 1A #status predicted <VP4>
F:246-491/Product: coat protein 1B #status predicted <VP5>
F:492-836/Product: coat protein 1C #status predicted <VP3>
F:837-980/Product: coat protein 1D #status predicted <VP1>
F:981-1108/Product: core protein 2A #status predicted <P2A>
F:1109-1438/Product: core protein 2B #status predicted <P2B>
F:1439-1496/Product: core protein 2C #status predicted <P2C>
F:1497-1519/Product: genome-linked protein VpG #status predicted <VPG>
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match      100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  KVNFPHGMDLEBIANSKD  20
|||||
Db      922  KVNFPHGMDLEBIANSKD  941

RESULT  5
GNMYSA
genome polypeptide - simian hepatitis A virus (strain AGM-27)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
C:Species: simian hepatitis A virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jun-2000
C:Accession: A30470; S04885; S03965
R:Tsarev, S.A.
submitted to JIPID, April 1991
A:Reference number: A30470
A:Accession: A30470
A:Molecule type: genomic RNA
A:Residues: 1-2230 <TSA>
A:Cross-references: GB:D00924; NID:g222597; PIDN:BAA00766.1; PID:g222598
R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.
J. Gen. Virol. 72, 1677-1683, 1991
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure
A:Reference number: J01080; MUID:91311420
A:Contents: annotation
```



A:Molecule type: DNA  
A:Residues: 1-1060 <STO>  
A:Cross-references: GB:chr\_IV; PIDN:AA37736.1; PID:g1208856; GSPDB:GN00022; CESP:CO1G5;  
C:Genetics:  
A:Gene: CO1G5.4  
A:Map position: 4

Query Match 45.2%; Score 47; DB 2; Length 1060;  
Best Local Similarity 36.8%; Pred. No. 22;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

OY 2 VNPPHGMLEETIANSKD 20  
DB 391 IGFSRGILDKHVAGDARD 409

RESULT 10  
T30996  
hypothetical protein CO1G5.4 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 22-Oct-1999 #sequence\_reviston 22-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30996  
R:Bradshaw, H.; Stellyes, L.  
submitted to the EMBL Data Library, August 1999  
A:Description: The sequence of C. elegans cosmid CO1G5.  
A:Reference number: Z20956  
A:Accession: T30996  
A>Status: preliminary; translated from GB/EMBL/DDbJ  
A:Molecule type: DNA  
A:Residues: 1-1079 <BRA>  
A:Cross-references: EMBL:U50068; PIDN:AA37736.2  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Map position: IV  
A:introns: 10/1; 31/1; 75/1; 108/1; 144/1; 927/2; 1056/2  
A:Note: CO1G5.4

Query Match 45.2%; Score 47; DB 2; Length 1079;  
Best Local Similarity 36.8%; Pred. No. 22;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

OY 2 VNPPHGMLEETIANSKD 20  
DB 410 IGFSRGILDKHVAGDARD 428

RESULT 11  
F82087  
deoxyribose-phosphate aldolase VC2350 [imported] - *Vibrio cholerae* (strain N16961 serog  
C:Species: *Vibrio cholerae*  
C>Date: 18-Aug-2000 #sequence\_reviston 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82087  
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolova, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833  
A:Accession: F82087  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <HEI>  
A:Cross-references: GB:AE004305; GB:AE003852; NID:g9656912; PIDN:AAF95493.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2350  
A:Map position: 1  
C:Superfamily: deoxyribose-phosphate aldolase

Query Match 44.2%; Score 46; DB 2; Length 259;

Best Local Similarity 52.9%; Pred. No. 6.2;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
OY 3 NPPHGMLEETIANSK 19  
DB 75 NPPHGMLEETIANSK 91

RESULT 12  
C89768  
conserved hypothetical protein SA0083 [imported] - *Staphylococcus aureus* (strain N315  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_reviston 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: C89768  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: C89768  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-444 <KUR>  
A:Cross-references: GB:BA000018; PID:g13700003; PIDN:BA841302.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0083

Query Match 44.2%; Score 46; DB 2; Length 444;  
Best Local Similarity 52.6%; Pred. No. 12;  
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 VNPPHGMLEETIANSKD 20  
DB 384 VNPPHGMLEETIANSKD 402

RESULT 13  
H64574  
DNA topoisomerase I - *Helicobacter pylori* (strain 26695)  
C:Species: *Helicobacter pylori*  
C>Date: 09-Aug-1997 #sequence\_reviston 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: H64574  
R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney,  
A.; Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,  
A.; title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
A:Reference number: A64520; MUID:97394467  
A:Accession: H64574  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-677 <TOM>  
A:Cross-references: GB:AE000559; GB:AE000511; NID:g2313536; PIDN:AAD07502.1; PID:g231  
C:Superfamily: DNA topoisomerase I

Query Match 44.2%; Score 46; DB 2; Length 677;  
Best Local Similarity 31.6%; Pred. No. 19;  
Matches 6; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIANSK 19  
DB 346 RITHPHALMDLEKVCSDAK 364

RESULT 14  
T08812  
Probable succinate--CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain - human (fragmen  
N; Alternate names: protein DKFZP586M203.1; succinyl-CoA synthetase (GDP-forming) bet

C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 05-May-2000  
C:Accession: T08812  
R:Ansong, W.; Winkler, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
Submitted to the Protein Sequence Database, May 1999  
A:Reference number: 216472  
A:Accession: T08812  
A:Molecule type: mRNA  
A:Residues: 1-195 <ANS>  
A:Cross-references: EMBL:AL050226  
A:Experimental source: adult uterus; clone DKFZP586M2023  
C:Genetics:  
A:Note: DKFZP586M2023.1  
C:Function:  
A:Description: catalyzes the formation of succinyl-CoA from succinate with concomitant H<sub>2</sub>O  
C:Superfamily: succinate--CoA ligase (ADP-forming) beta chain  
C:Keywords: acid-thiol ligase; coenzyme A; mitochondrion

Query Match 43.3%; Score 45; DB 2; Length 195;  
Best Local Similarity 42.1%; Pred. No. 6.5;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 2 VNPPHGMLEIEIANSKD 20  
| | | | | : | : | : | : | : | :  
Db 65 VSPGGVDIEVAASNP 83

RESULT 15  
F96582  
hypothetical protein F1511.25 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: F96582  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: F96582  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-383 <STO>  
A:Cross-references: GB:AE005173; NID:94587558; PIDN:AAD25789.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F1511.25  
A:Map position: 1

Query Match 43.3%; Score 45; DB 2; Length 383;  
Best Local Similarity 44.4%; Pred. No. 14;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 3 NPPHGMLEIEIANSKD 20  
| | | | | : | : | : | : | : | :  
Db 209 DYPGLDIEIKLPGRD 226

RESULT 16  
A44529  
succinate--CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 30-Sep-1993 #sequence\_revision 07-Apr-1994 #text\_change 05-May-2000  
C:Accession: A44529  
R:Baily, D.L.; Molodko, W.T.; Bridger, W.A.  
Protein Sci. 2, 1255-1262, 1993  
A:Title: Cloning, characterization and expression of the beta subunit of pig heart succi

A:Reference number: A44529; MUID:94004462

A:Accession: A44529  
A:Molecule type: mRNA  
A:Residues: 1-417 <BAI>  
A:Cross-references: GB:I06944  
C:Genetics:  
A:Start codon: CTG  
C:Function:  
A:Description: catalyzes the formation of succinyl-CoA from succinate with concomitan  
C:Superfamily: succinate--CoA ligase (ADP-forming) beta chain  
C:Keywords: acid-thiol ligase; coenzyme A; heterodimer; mitochondrion  
F:1-23/Domain: transit peptide (mitochondrion) (fragment) #status predicted <TNP>  
F:24-417/Product: succinate--CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain #status

Query Match 43.3%; Score 45; DB 2; Length 417;  
Best Local Similarity 42.1%; Pred. No. 16;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 2 VNPPHGMLEIEIANSKD 20  
| | | | | : | : | : | : | : | :  
Db 153 VSPGGVDIEVAASNP 171

RESULT 17  
B72685  
hypothetical protein APE0898 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: B72685  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339  
A:Accession: B72685  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-852 <KAW>  
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79882.1; PID:95104567  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0898  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0898

Query Match 43.3%; Score 45; DB 2; Length 852;  
Best Local Similarity 41.2%; Pred. No. 36;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 VNPPHGMLEIEIANS 18  
| | | | | : | : | : | : | : | :  
Db 355 VNGPYGLIDISDIYVDS 371

RESULT 18  
T19011  
hypothetical protein C06C6.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19011  
R:McMurray, A.  
submitted to the EMBL data library, March 1997  
A:Reference number: Z19059  
A:Accession: T19011  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-935 <WIL>  
A:Cross-references: EMBL:Z93374; PIDN:CAB07557.1; GSPDB:GN00023; CESP:C06C6.7  
C:Experimental source: clone C06C6  
C:Genetics:  
A:Gene: CESP:C06C6.7  
A:Map position: 5





A:Residues: 298-312,'L',314-500 <CHA>  
C:Genetics:  
A:Gene: SGD:ANP1  
A:Cross-references: SGD:S0000762; MIPS:YEL036c  
A:Map position: 5L

Query Match 42.3%; Score 44; DB 2; Length 500;  
Best Local Similarity 31.6%; Pred. No. 29;  
Matches 6; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNPFGMLDLEEFIANSKD 20  
:||:||||:|  
DB 106 MYPFHNLIDLSFLVSDSD 124

## RESULT 23

hypothetical protein ydCC - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: A69773

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.: Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
lech, J.; Harwood, C.R.; Hanaud, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69773  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-373 <RUN>

A:Cross-references: GB:299106; GB:AL009126; NID:g2632653; PIDN:CAB12270.1; PID:el182429;

A:Experimental source: strain 168

C:Genetics:

A:Gene: ydCC  
Query Match 41.8%; Score 43.5; DB 2; Length 373;  
Best Local Similarity 47.6%; Pred. No. 25;  
Matches 10; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 KVNPHG-MLDLEEFIANSKD 20  
||:||||:|  
DB 186 KTNQHNKMLPTQETPFKKD 206

## RESULT 24

E81659

conserved hypothetical protein TC0838 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 18-Aug-2000

C:Accession: E81659

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: E81659

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-140 <PRT>

A:Cross-references: GB:AE002350; GB:AE002160; NID:g7190860; PIDN:AAF39638.1; PID:g719086

A:Experimental source: strain Nigg (Mopn)  
C:Genetics:  
A:Gene: TC0838  
C:Superfamily: conserved hypothetical protein CP0076

Query Match 41.3%; Score 43; DB 2; Length 140;  
Best Local Similarity 46.7%; Pred. No. 9.6;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 FPHGMLDLEEFIANNS 18  
|||:||||:|  
DB 23 FPEGFLDISDVLAARS 37

## RESULT 25

G71501

hypothetical protein CF550 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 18-Aug-2000

C:Accession: G71501

R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: G71501

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-141 <ARN>

A:Cross-references: GB:AE001325; GB:AE001273; NID:g3328980; PIDN:AAC68152.1; PID:g332

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: CF550

C:Superfamily: conserved hypothetical protein CP0076

Query Match 41.3%; Score 43; DB 2; Length 141;  
Best Local Similarity 46.7%; Pred. No. 9.7;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 FPHGMLDLEEFIANNS 18  
|||:||||:|  
DB 21 FPEGFLDISDVLAARS 35

Search completed: June 16, 2002, 00:05:22  
Job time: 796 sec

---



DR Pfam; PF00680; RNA\_dep\_RNA\_pol: 1.  
DR Pfam; PF00910; RNA\_helicase: 1.  
KW Polypeptide; Coat protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 CORE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BE75 CRC64;

Query Match 100.0%; Score 104; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 1,1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNFPFGMLDEEIANSKD 20  
DB 922 KVNFPFGMLDEEIANSKD 941  
|||||

RESULT 2  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91162758; PubMed-1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M., Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination."  
RT J. Virol. 65:2056-2065(1991).  
RL -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M59808: AAA45467.1; -.  
CC MEROPS: C03.005; -.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol: 1.  
DR Pfam: PF00910; RNA\_helicase: 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).

FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A6396C8D5B CRC64;

Query Match 100.0%; Score 104; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 1,1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNFPFGMLDEEIANSKD 20  
DB 922 KVNFPFGMLDEEIANSKD 941  
|||||

RESULT 3  
POLG\_HPAV8 STANDARD; PRT; 2227 AA.  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12098;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATTENUATED;  
RX MEDLINE-87175701; PubMed-3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Felstone S.M., Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with wild-type virus."  
RT Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RL [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=ATTENUATED;  
RX MEDLINE-85166289; PubMed-2984684;  
RA Baroudy B.M., Ticehurst J.R., Mele T.A., Malzel J.V. Jr., Purcell R.H., Felstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA polymerase."  
RT Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
RL -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT SHOWN.  
CC -----

```

CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; M14114; AAA45475.1; -
DR      EMBL; M14707; AAA45465.1; -
DR      EMBL; M14707; AAA45466.1; ALT_INIT.
DR      PIR; A25981; GNNYHM.
DR      PIR; A25914; GNNYMK.
DR      PIR; A03905; A03905.
DR      MEROPS; C03.005; -
DR      InterPro; IPR000605; RNA_helicase.
DR      InterPro; IPR01205; RNA_pol_P3D.
DR      Pfam; PF00680; RNA_dep_RNA_pol_1.
DR      Pfam; PF00910; RNA_helicase; 1.
KW      Polypeptide; Coat protein; Core protein; Transferrase;
KW      RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT      CHAIN 1 23
FT      CHAIN 24 245
FT      CHAIN 246 491
FT      CHAIN 492 836
FT      CHAIN 837 980
FT      CHAIN 981 1087
FT      CHAIN 1088 1422
FT      CHAIN 1423 1496
FT      CHAIN 1497 1519
FT      CHAIN 1520 1738
FT      CHAIN 1739 2227
FT      CHAIN 77 77
FT      VARIANT 764 764
FT      VARIANT 821 821
FT      VARIANT 1052 1052
FT      VARIANT 1062 1062
FT      VARIANT 1118 1118
FT      VARIANT 1151 1151
FT      VARIANT 1163 1163
FT      VARIANT 1277 1277
FT      VARIANT 1500 1805
FT      VARIANT 1805 1805
FT      VARIANT 1930 1930
SQ      SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDEETIAANSKD 20
Db 922 KVNFPHGMLDEETIAANSKD 941
|||||
|||||

RESULT 4
POLG_HPAVL STANDARD; PRT; 2227 AA.
AC P06441;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Coat proteins VP1 to VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (Ec 2.7.7.48)].
OS Hepatitis A virus (strain LA).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_Taxid=12099;
FM [1]
RP SEQUENCE FROM N.A.

```

```

RA MEDLINE=65190549; PubMed-2986127
RX Najarian R., Caput D., Gee W.M., Potter S.J., Renard A.,
RA Merryweather J., van Nest G., Dina D.;
RL "Primary structure and gene organization of human hepatitis A virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
CC -1 SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1 PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1 SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; K02990; AAAA5472.1; .
DR PIR; A03903; GNMYHR.
DR MEROPS; C03.005; -.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIV) .
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB) .
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC) .
FT CHAIN 492 836 COAT PROTEIN VP1 (PID) .
FT CHAIN 837 980 CORE PROTEIN P2A.
FT CHAIN 981 1076 CORE PROTEIN P2B.
FT CHAIN 1077 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1484 PROBABLE PROTEIN P3A.
FT CHAIN 1485 1507 PROBABLE PROTEIN P3B.
FT CHAIN 1508 1678 PROBABLE PROTEIN P3C.
FT CHAIN 1679 2227 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354BCD2799C CRC64;

Query Match 100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1,1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPGHMLDEETIANSKD 20
Db 922 KVNFPGHMLDEETIANSKD 941
|||||
POLG_HPAVM STANDARD; PRF: 2227 AA.
AC PI3901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
OC Q81090; Q81091; Q81092; Q81093;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
P3D (EC 2.7.7.48)].
DE Hepatitis A virus (strain MBV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=68045071; PubMed=2823500;
RA Paul A.V., Tada H., der Helm K., Wissel T., Klein R., Wimmer E.,
RA Deinhardt F.;
RT "The entire nucleotide sequence of the genome of human hepatitis A
virus (isolate MBV).";
```

Virus Res. 8:153-171(1987).  
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
-1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
-1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
WHO HAD BEEN INFECTED IN THE AFRICAN PEPTIDASE FAMILY C3.  
-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
-----  
DR EMBL: M20273; AAA5474.1; -  
DR PIR: J50303; GNNYHB.  
DR MEROPS: C03.005; -  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol.1.  
DR Polyprotein: Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
SO SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
  
Query Match 100.0%; Score 104; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KVNFPFGMDLEIANSKD 20  
DB 922 KVNFPFGMDLEIANSKD 941  
  
RESULT 6  
POLG\_HPAVS STANDARD: PRT: 2230 AA.  
AC P14553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311420; PubMed=1649901;  
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,  
Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RX SEQUENCE OF 1750-2164 FROM N.A.  
RP MEDLINE=89232168; PubMed=2541023;  
RX Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,

RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
and simian hepatitis A viruses.";  
RL FEBS Lett. 247:425-428(1989).  
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
VP3, AND VP4.  
-1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
-----  
DR EMBL: D00924; BAA00766.1; -  
DR PIR: A30470; GNNYSA.  
DR PIR: S04885; S04885.  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol.1.  
DR Pfam: PF00910; RNA\_helicase.1.  
DR Polyprotein: Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 27  
FT CHAIN 28 249  
FT CHAIN 250 495  
FT CHAIN 496 795  
FT CHAIN 796 984  
FT CHAIN 985 1091  
FT CHAIN 1092 1426  
FT CHAIN 1427 1458  
FT CHAIN 1459 1521  
FT CHAIN 1522 1741  
FT CHAIN 1742 2230  
SO SEQUENCE 2230 AA; 251296 MW; B7B3330E324E1F19 CRC64;  
  
Query Match 97.1%; Score 101; DB 1; Length 2230;  
Best Local Similarity 95.0%; Pred. No. 3.8e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KVNFPFGMDLEIANSKD 20  
DB 926 KVNFPFGMDLEIANSKD 945  
  
RESULT 7  
POLG\_HPAV2 STANDARD: PRT: 2226 AA.  
AC P26560;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
variants arising during persistent infection: evidence for genetic



```

RL EMBO J. 14:4156-4167(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HB8 / ATCC 27634;
RX MEDLINE-98149692; PubMed-9490048;
RA Mazuric M.-H., Keith G., Logan D., Kreutzer R., Giege R., Kern D.;
RT "Glycyl tRNA synthetase from Thermus thermophilus -- wide structural
divergence with other prokaryotic glycyl-tRNA synthetases and
functional inter-relation with prokaryotic and eukaryotic glycylation
systems."
RT Eur. J. Biochem. 251:744-757(1998).
CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ222643; CA10903.1; -.
CC PDB: 1AT1; 07-JUL-97.
CC PDB: 1B76; 28-JAN-99.
CC PDB: 1GGM; 28-JAN-99.
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR004154; HGTP_antlicodon.
DR InterPro: IPR002314; tRNA-synt_2b.
DR InterPro: IPR002315; tRNA-synt_gly.
DR Pfam: PF03129; HGTP_antlicodon; 1.
DR Pfam: PF00587; tRNA-synt_2b; 1.
DR PRINTS: PR01043; TRNASYNTHGLY.
DR PROSITE: PS00179; AA_TRNA_LIGASE_II_1; 1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; PALSE.NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW 3D-structure.
FT INIT_MER 0 0
FT CONFLICT 1 1 A->P (IN REF. 2).
FT CONFLICT 11 11 YRAME->TGPMR (IN REF. 2).
FT CONFLICT 191 199 OGIFVFNK->RASSSTSR (IN REF. 2).
FT CONFLICT 215 219 IGNKF->SARPS (IN REF. 2).
FT CONFLICT 266 266 E->R (IN REF. 2).
FT CONFLICT 283 284 SS->EL (IN REF. 2).
FT CONFLICT 302 303 SL->LE (IN REF. 2).
FT CONFLICT 310 310 Q->N (IN REF. 2).
SQ SEQUENCE 505 AA; 58081 MW; EC1C8D5388AB7570 CRC64;

Query Match
Best local Similarity 45.2%; Score 47; DB 1; Length 505;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 FPHGMIDLEIEIA 15
Db 298 FPHGSLEIEGIA 309

RESULT 10
DEOC_VIRCH
ID DEOC_VIRCH STANDARD; PRT; 259 AA.
AC 09KPL7;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyribaldolase)
DE (Deoxyribaldolase).
DEOC OR VC2350.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

```

```

OX NCBI_TaxId-666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE-20406833; PubMed-10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unanue L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RT Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
glyceraldehyde 3-phosphate + acetaldehyde.
CC -1- PATHWAY: NUCLEOTIDE AND DEOXYRIBONUCLEOTIDE CATABOLISM.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAP FAMILY OF ALDOLASES.
CC DEOC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004305; AAF95493.1; -.
CC TIGR: VC2350; -.
DR InterPro: IPR002915; Deoc.
DR Pfam: PF01791; Deoc; 1.
KW Lyase; Schiff base; Complete proteome.
FT BINDING 166 SCHIFF-BASE (BY SIMILARITY).
SQ SEQUENCE 259 AA; 27959 MW; 2B76B07AC6527C73 CRC64;

Query Match
Best local Similarity 44.2%; Score 46; DB 1; Length 259;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 NEPHGMIDLEIEIANSK 19
Db 75 NEPHGMIDLEIEIAVERK 91

RESULT 11
SUChB_PIG
ID SUChB_PIG STANDARD; PRT; 417 AA.
AC P53590; Q95279;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Succinyl-CoA ligase (GDP-forming) beta-chain, mitochondrial precursor
DE (EC 6.2.1.4) (Succinyl-CoA synthetase, beta chain) (SCS-beta).
GN SUCCLG2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId-9823;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-41.
RC TISSUE-Heart;
RX MEDLINE-94004462; PubMed-8401211;
RA Bailey D.L., Molodko W.T., Bridger W.A.;
RT "Cloning, characterization, and expression of the beta subunit of pig
heart succinyl-CoA synthetase."
RL Protein Sci. 2:1255-1262(1993).
RN [2]
RP SEQUENCE OF 1-82 FROM N.A.
RC TISSUE-Small intestine;

```



```

RX MEDLINE-96327607; PubMed-8672129;
RA Manteo A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RL Mamm. Genome 7:509-517(1996).
CC -1- CATALYTIC ACTIVITY: GTP + succinate + CoA = GDP + succinyl-CoA +
CC phosphate.
CC -1- PATHWAY: SUBSTRATE LEVEL PHOSPHORYLATION STEP OF THE TRICARBOXYLIC
CC ACID CYCLE.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: TO OTHER BETA SUBUNITS OF SUCCINYL-COA SYNTHETASE,
CC OF MALATE-COA LIGASE AND TO ATP CITRATE-LYASE.
CC -1- CAUTION: THE AUTHORS INDICATE THAT THE CODON CTG CODING FOR
CC LEU-1 MAY SERVE AS THE INITIATOR.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: L06944; AAA31120.1; -.
DR EMBL: Z81187; CAB03559.1; -.
DR HSSP: P07460; ISCU.
DR InterPro: IPR003135; ATP-grasp.
DR InterPro: IPR000303; CoA_ligase.
DR Pfam: PF02222; ATP-grasp; 1.
DR Pfam: PF00549; Ligase-CoA; 1.
DR PROSITE: PS01217; SUCCINYL-CoA_Lig_3; 1.
KW Ligase; Glycolysis; Tricarboxylic acid cycle; Mitochondrion;
KW Transit peptide.
FT TRANSIT 1 22 MITOCHONDRION.
FT CHAIN 23 417 SUCCINYL-COA LIGASE [GDP-FORMING] BETA-
FT FT CHAIN.
SQ SEQUENCE 417 AA; 45131 MW; F09DDCC29B077CC CRC64;

Query Match 43.3%; Score 45; DB 1; Length 417;
Best Local Similarity 42.1%; Pred. No. 3.6;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 2 VNPFGMDLEBIANSKD 20
Db 153 VGSPOGVDIEVANSNP 171

RESULT 12
ANPL_YEAST
ID ANPL_YEAST STANDARD; PRT; 500 AA.
AC P32629;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Amonitrophenyl propanediol resistance protein.
GN ANP1 OR GEM3 OR YEL036C OR YSGP-ORF28.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-B-6441;
RX MEDLINE-94016558; PubMed-8411151;
RA Melnick L., Sherman F.;
RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
RT of Saccharomyces cerevisiae share a common ancestry.";
RL J. Mol. Biol. 233:372-386(1993).
[2]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,

```

```

RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[4]
RP CHARACTERIZATION.
RX MEDLINE-95045382; PubMed-7957057;
RA Chapman R.E., Munro S.;
RT "The functioning of the yeast Golgi apparatus requires an ER protein
RT encoded by ANP1, a member of a new family of genes affecting the
RT secretory pathway.";
RL EMBL J. 13:4896-4907(1994).
CC -1- FUNCTION: INVOLVED IN THE ORGANIZATION OF THE SECRETORY PATHWAY.
CC REQUIRED TO MAINTAIN A FUNCTIONAL GOLGI APPARATUS.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -1- SIMILARITY: BELONGS TO THE ANP1 / WMN9 / VAN1 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: S65964; AAD13971.1; -.
DR EMBL: L22171; AAA34426.1; -.
DR EMBL: S66114; AAB28440.1; -.
DR EMBL: L22173; AAA34937.1; -.
DR EMBL: U18779; AAB65006.1; -.
DR PIR: S30846; S30846.
DR PIR: S38346; S38346.
DR SGD: S0000762; ANP1.
KW Glycoprotein; Transmembrane; Signal-anchor; Golgi stack.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 28 500 LUMENAL (POTENTIAL).
FT DOMAIN 446 467 POLY-GLN.
FT CONFLICT 220 224 HHDKD -> OSQGN (IN REF. 1).
FT CONFLICT 313 313 F -> L (IN REF. 1).
FT CONFLICT 472 500 PGGKRLDNDKKKKHKKHREYLDPPDRN -> RRGMLMT
FT FT TTPRRKNIKKH (IN REF. 1).
SQ SEQUENCE 500 AA; 58182 MW; 845B393CE54BCD14 CRC64;

Query Match 42.3%; Score 44; DB 1; Length 500;
Best Local Similarity 31.6%; Pred. No. 6.7;
Matches 6; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 2 VNPFGMDLEBIANSKD 20
Db 106 MTPPHNIDLSFVSDSSD 124

RESULT 13
SRG1_HUMAN
ID SRG1_HUMAN STANDARD; PRT; 98 AA.
AC O75711;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Scrapie-responsive protein 1 precursor (SCRG-1).
GN SCRG1.
OS Homo sapiens (Human).

```





Best Local Similarity 63.6%; Pred. NO. 5.8;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 FPHGMDLEE1 14  
DB 146 FPGGLDFOEI 156

# RESULT 18

YC06\_YEAST STANDARD; PRT; 290 AA.  
AC P25617;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE Hypothetical 33.6 kDa protein in POL4-SRD1 intergenic region.  
GN YCR016W OR YCR16W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hatat D., Jacq C., Perea J., Shu Y.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: X59720; CAA42333.1; -  
DR PIR: S19426.  
DR SGD: S0000609; YCR016W.  
KW Hypothetical protein.  
SQ SEQUENCE 290 AA; 33621 MW; 72458EF03717341 CRC64;

Query Match Best Local Similarity 40.4%; Score 42; DB 1; Length 290;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 FPHGMDLEE1ANSKD 20  
DB 116 FPGGLEESYKENYGD 132

RESULT 19  
YB71\_HAEIN STANDARD; PRT; 193 AA.  
AC P44339;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Putative anthranilate synthase component II (EC 4.1.3.27) (Glutamine  
DE amido-transferase).  
GN H11171.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SRRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).

CC -1- CATALYTIC ACTIVITY: Chorismate + L-glutamine -> anthranilate +  
CC pyruvate + L-glutamate.  
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
CC -1- SUBUNIT: Tetramer of two components I and two components II.  
CC -1- SIMILARITY: TO OTHER TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAINS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: U32797; AAC22824.1; -  
DR HSSP: 006129; IQDL.  
DR TIGR: H11171; -  
DR InterPro: IPR000991; GATase\_1.  
DR Pfam: PF00117; GATase; 1.  
DR PRINTS: PR00096; GATASE.  
DR PROSITE: PS00442; GATASE\_TYPE\_1; 1.  
KW Hypothetical protein: Tryptophan biosynthesis; Lyase;  
KW glutamine amidotransferase; Complete proteome.  
FT ACT\_SITE 78 78 BY SIMILARITY.  
FT ACT\_SITE 168 168 BY SIMILARITY.  
FT ACT\_SITE 170 170 BY SIMILARITY.  
SQ SEQUENCE 193 AA; 22300 MW; 2C6468C44DE56999 CRC64;

Query Match Best Local Similarity 39.4%; Score 41; DB 1; Length 193;  
Matches 6; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

OY 1 KVNPHGMDLEE1ANSKD 20  
DB 22 KLVNPIVUNVEDIKENTAE 41

RESULT 20  
Y363\_RICPR STANDARD; PRT; 218 AA.  
AC Q9ZDG6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein RP363.  
GN RP363.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SRRAIN-MADRID E;  
RX MEDLINE=99039459; PubMed=9823893;  
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,  
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria.";  
RL Nature 396:133-140(1998).

CC -1- SIMILARITY: SOME, TO R.PROWAZEKII RP364.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AJ235271; CAA14822.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 218 AA; 25780 MW; ZDBBD7DB938F954 CRC64;

Query Match 39.4%; Score 41; DB 1; Length 218;  
Best Local Similarity 38.9%; Pred. No. 8.8;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 KVNPHGMLDELEITANS 18  
DB 49 KINPSPKILELYKLILNN 66

RESULT 21  
ID NIFS\_ANAA2 STANDARD; PRT; 400 AA.  
AC 043884;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cysteine desulfurase (EC 4.4.1.-) (Nitrogenase metalloclusters  
biosynthesis protein nifs).  
GN NIFS.  
OS Anabaena azollae.  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
OX NCBI\_TaxID=1164;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1A;  
RA MEDLINE=96118705; Pubmed=7496536;  
RA Jackson D.M., Mulligan M.E.;  
RT "Characterization of a nitrogen-fixation (nif) gene cluster from  
RT Anabaena azollae 1a shows that closely related cyanobacteria have  
RT highly variable but structured intergenic regions.";  
RL Microbiology 141:2235-2244(1995).  
CC -1- FUNCTION: CATALYZES THE REMOVAL OF ELEMENTAL SULFUR ATOMS FROM  
CC CYSTEINE TO PRODUCE ALANINE. SEEMS TO PARTICIPATE IN THE  
CC BIOSYNTHESIS OF THE NITROGENASE METALLOCLUSTERS BY PROVIDING THE  
CC INORGANIC SULFUR REQUIRED FOR THE FE-S CORE FORMATION (BY  
CC SIMILARITY).  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES. NIFS/ISCS SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L34879; AAA87249.1; -  
DR InterPro: IPR000192; AminoTransf\_class.V.  
DR Pfam: PF00266; aminoTran\_5.1.  
DR PROSITE: PS00595; AA\_TRANSF\_CLASS\_5; 1.  
KW Nitrogen fixation; Lyase; Pyridoxal phosphate.  
FT BINDING 201 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT ACT\_SITE 324 324 BY SIMILARITY.  
SQ SEQUENCE 400 AA; 43685 MW; 8DEF6B713C5C3A6E CRC64;

Query Match 39.4%; Score 41; DB 1; Length 400;  
Best Local Similarity 58.3%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 HGMLDLEITAN 17  
DB 124 HQGLDLEITANS 135

RESULT 22  
ID NIFS\_ANASP STANDARD; PRT; 400 AA.  
AC P12623;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Cysteine desulfurase (EC 4.4.1.-) (Nitrogenase metalloclusters  
biosynthesis protein nifs).  
GN NIFS OR AL1457.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90037054; Pubmed=2553733;  
RA Mulligan M.E., Haselkorn R.;  
RT "Nitrogen fixation (nif) genes of the cyanobacterium Anabaena species  
RT strain PCC 7120. The nifB-fixN-nifs-nifU operon.";  
RL J. Biol. Chem. 264:19200-19207(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; Pubmed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
RN [3]  
RP SEQUENCE OF 1-7 FROM N.A.  
RX MEDLINE=88314954; Pubmed=2842320;  
RA Mulligan M.E., Bulkema W.J., Haselkorn R.;  
RT "Bacterial-type ferredoxin genes in the nitrogen fixation regions of  
RT the cyanobacterium Anabaena sp. strain PCC 7120 and Rhizobium  
RT meliloti.";  
RL J. Bacteriol. 170:4406-4410(1988).  
CC -1- FUNCTION: CATALYZES THE REMOVAL OF ELEMENTAL SULFUR ATOMS FROM  
CC CYSTEINE TO PRODUCE ALANINE. SEEMS TO PARTICIPATE IN THE  
CC BIOSYNTHESIS OF THE NITROGENASE METALLOCLUSTERS BY PROVIDING THE  
CC INORGANIC SULFUR REQUIRED FOR THE FE-S CORE FORMATION (BY  
CC SIMILARITY).  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES. NIFS/ISCS SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: J05111; AAA22006.1; -  
DR EMBL: AP003585; BAB73413.1; -  
DR PIR: B32361; B32361.  
DR PIR: C34443; C34443.  
DR InterPro: IPR000192; AminoTransf\_class.V.  
DR Pfam: PF00266; aminoTran\_5.1.  
DR PROSITE: PS00595; AA\_TRANSF\_CLASS\_5; 1.  
KW Nitrogen fixation; Lyase; Pyridoxal phosphate; Complete proteome.  
FT BINDING 201 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT ACT\_SITE 324 324 BY SIMILARITY.  
FT CONFLICT 88 88 E -> A (IN REF. 1).



OS	Rickettsia prowazekii.
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC	Rickettsiaceae; Rickettsiae; Rickettsia.
OX	NCBI_TaxID:782;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN-MADRID E.
RX	MEDLINE=97419517; PubMed=9274032;
RA	Andersson J.O., Andersson S.G.E.:
RT	"Genomic rearrangements during evolution of the obligate
RT	intracellular parasite Rickettsia prowazekii as inferred from an
RT	analysis of 52015 bp nucleotide sequence.";
RL	Microbiology 143:2783-2795(1997).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-MADRID E.
RX	MEDLINE=99039499; PubMed=9823893;
RA	Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA	Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA	Eriksson A.-S., Winkler H.H., Kurland C.G.:
RT	"The genome sequence of Rickettsia prowazekii and the origin of
RT	mitochondria.";
RL	Nature 396:133-140(1998).
CC	-1 CATALYTIC ACTIVITY: UDP-glucose + 2 NAD(+) + H(2)O = UDP-
CC	glucuronate + 2 NADH.
CC	-1 SIMILARITY: BELONGS TO THE UDP-GLUCOSE/GDP-MANNOSE DEHYDROGENASES
CC	FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: Y11785; CAA72478.1; -
DR	EMBL: AJ235273; CAA15205.1; -
DR	InterPro: IPR001732; UDPG_MGDP_dh.
DR	Pfam: PF00984; UDPG_MGDP_dh; 1.
KW	Oxidoreductase; NAD; Complete proteome.
FT	NP_BIND 2 19 NAD (POTENTIAL).
FT	ACT_SITE 260 260 BY SIMILARITY.
SO	SEQUENCE 434 AA; 48226 MW; 15991BA7112D5036 CRC64;

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:53 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-44  
Perfect score: 104  
Sequence: 1 KVNPPHGMLEIEIAANSKD 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-misc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteriaph:\*  
17: SP-archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	251	12 09ENN1	Q9enn1 hepatitis a
2	104	100.0	251	12 09ENN9	Q9enn9 hepatitis a
3	104	100.0	251	12 09ENN7	Q9enn7 hepatitis a
4	104	100.0	251	12 09ENN6	Q9enn6 hepatitis a
5	104	100.0	251	12 09ENN5	Q9enn5 hepatitis a
6	104	100.0	251	12 09ENN4	Q9enn4 hepatitis a
7	104	100.0	251	12 09ENN1	Q9enn1 hepatitis a
8	104	100.0	251	12 09ENN9	Q9enn9 hepatitis a
9	104	100.0	251	12 09ENN7	Q9enn7 hepatitis a
10	104	100.0	251	12 09ENN5	Q9enn5 hepatitis a
11	104	100.0	251	12 09ENN4	Q9enn4 hepatitis a
12	104	100.0	251	12 09ENN1	Q9enn1 hepatitis a
13	104	100.0	251	12 09ENN9	Q9enn9 hepatitis a
14	104	100.0	251	12 09ENN7	Q9enn7 hepatitis a
15	104	100.0	251	12 09ENN5	Q9enn5 hepatitis a
16	104	100.0	251	12 09ENN4	Q9enn4 hepatitis a

## ALIGNMENTS

17	104	100.0	251	12 09ENN2	Q9enn2 hepatitis a
18	104	100.0	1124	12 084780	084780 hepatitis a
19	104	100.0	1161	12 005794	005794 hepatitis a
20	104	100.0	2216	12 09WMA2	09WMA2 hepatitis a
21	104	100.0	2218	12 067824	067824 hepatitis a
22	104	100.0	2225	12 09DLJ32	09DLJ32 hepatitis a
23	104	100.0	2227	12 067825	067825 hepatitis a
24	104	100.0	2227	12 067826	067826 hepatitis a
25	104	100.0	2227	12 09WMA4	09WMA4 hepatitis a
26	104	100.0	2227	12 09WMA3	09WMA3 hepatitis a
27	104	100.0	2227	12 09WMA1	09WMA1 hepatitis a
28	104	100.0	2227	12 09WMA0	09WMA0 hepatitis a
29	104	100.0	2227	12 091FHS	091FHS hepatitis a
30	98	94.2	251	12 09ENR0	09ENR0 hepatitis a
31	98	94.2	251	12 09ENR8	09ENR8 hepatitis a
32	98	94.2	251	12 09ENR3	09ENR3 hepatitis a
33	98	94.2	251	12 09ENR0	09ENR0 hepatitis a
34	98	94.2	251	12 09ENR8	09ENR8 hepatitis a
35	98	94.2	251	12 09ENR3	09ENR3 hepatitis a
36	98	94.2	251	12 09ENR6	09ENR6 hepatitis a
37	98	94.2	251	12 09ENP0	09ENP0 hepatitis a
38	98	94.2	251	12 09ENN8	09ENN8 hepatitis a
39	98	94.2	251	12 09ENN3	09ENN3 hepatitis a
40	97	93.3	184	12 087092	087092 hepatitis a
41	95	91.3	2218	12 067817	067817 hepatitis a
42	94	90.4	251	12 09ENQ2	09ENQ2 hepatitis a
43	94	90.4	2227	12 09WMP9	09WMP9 hepatitis a
44	93	89.4	251	12 09ENR4	09ENR4 hepatitis a
45	93	89.4	251	12 09ENN9	09ENN9 hepatitis a
46	47	45.2	404	11 092218	092218 hepatitis a
47	47	45.2	1079	5 017566	017566 mus muscu
48	46	44.2	39	5 09NC67	09NC67 caenorhabd
49	46	44.2	444	16 099XC5	099XC5 staphylococ
50	46	44.2	677	16 025188	025188 helicobacte
51	45	43.3	195	4 09T436	09T436 homo sapien
52	45	43.3	383	10 09SYH3	09SYH3 arbidopsi
53	45	43.3	432	4 096199	096199 homo sapien
54	45	43.3	852	17 09YDL5	09YDL5 aeropyrum p
55	45	43.3	935	5 062033	062033 caenorhabd
56	45	43.3	1213	5 09XTC7	09XTC7 caenorhabd
57	45	43.3	3199	3 0960W4	0960W4 usliiag ma
58	44.5	42.8	218	14 0910U9	0910U9 plasmodi psb
59	44	42.3	223	3 013832	013832 schizosacch
60	44	42.3	344	16 0921C8	0921C8 rickettsia

RESULT 1 PRELIMINARY: PRT: 251 AA.

Q9ENN1: 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
NCBI\_TaxID=12092;  
[1]  
RC STRAIN=AL;  
RA Fujiiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047652; BAB12160.1; --  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA: 28749 MW: 58A520D873893445 CRC64;

```
Query Match          100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIAANSKD 20
    |||||||
Db 86 KVNPPHGMLEETIAANSKD 105

RESULT 2
O9EN09 PRELIMINARY; FRT; 251 AA.
AC O9EN09;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A159;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047654; BAB12162.1; -.
FT NON_TER 1 251
FT SEQUENCE 251 AA; 28720 MW; C334248282F19CA CRC64;

Query Match          100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIAANSKD 20
    |||||||
Db 86 KVNPPHGMLEETIAANSKD 105

RESULT 3
O9EN07 PRELIMINARY; PRT; 251 AA.
AC O9EN07;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A161;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047656; BAB12164.1; -.
FT NON_TER 1 251
FT SEQUENCE 251 AA; 28766 MW; A0EC02E3609C4D5B CRC64;

Query Match          100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIAANSKD 20
    |||||||
Db 86 KVNPPHGMLEETIAANSKD 105
```

```
RESULT 4
O9EN06 PRELIMINARY; PRT; 251 AA.
AC O9EN06;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A162;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047657; BAB12165.1; -.
FT NON_TER 1 251
FT SEQUENCE 251 AA; 28699 MW; 56AA86DF60F6FD3B CRC64;

Query Match          100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIAANSKD 20
    |||||||
Db 86 KVNPPHGMLEETIAANSKD 105

RESULT 5
O9EN05 PRELIMINARY; PRT; 251 AA.
AC O9EN05;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A20;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047658; BAB12166.1; -.
FT NON_TER 1 251
FT SEQUENCE 251 AA; 28720 MW; C334248282F19CA CRC64;

Query Match          100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPPHGMLEETIAANSKD 20
    |||||||
Db 86 KVNPPHGMLEETIAANSKD 105

RESULT 6
O9EN04 PRELIMINARY; PRT; 251 AA.
AC O9EN04;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
```

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus  
OX NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A201;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047659; BAB12167.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMDLEIANSKD 20  
DB 86 KVNPHGMDLEIANSKD 105

## RESULT 7

OGENO1 PRELIMINARY; PRT; 251 AA.  
AC OGENO1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A206;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047662; BAB12170.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28699 MW; 8EADAE7E2754C37 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMDLEIANSKD 20  
DB 86 KVNPHGMDLEIANSKD 105

## RESULT 8

OGENP9 PRELIMINARY; PRT; 251 AA.  
AC OGENP9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID-12092;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A303;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047664; BAB12172.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28752 MW; 7215A28AD2CA5C1A CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMDLEIANSKD 20  
DB 86 KVNPHGMDLEIANSKD 105

## RESULT 9

OGENP7 PRELIMINARY; PRT; 251 AA.  
AC OGENP7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A306;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047666; BAB12174.1; -.  
FT NON\_TER 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMDLEIANSKD 20  
DB 86 KVNPHGMDLEIANSKD 105

## RESULT 10

OGENP5 PRELIMINARY; PRT; 251 AA.  
AC OGENP5;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID-12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A407;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047668; BAB12176.1; -.

FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28614 MW; 8334EF19C757A6D CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPPHGMLDLEETIAANSKD 20  
DB 86 KVNPPHGMLDLEETIAANSKD 105

## RESULT 11

ID Q9ENP2 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP2;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC SEQUENCE FROM N.A.  
RA STRAIN-A5;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047671; BAB12179.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28622 MW; 355C3CD146D39D02 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPPHGMLDLEETIAANSKD 20  
DB 86 KVNPPHGMLDLEETIAANSKD 105

## RESULT 12

ID Q9ENP1 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN-A503;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047672; BAB12180.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28663 MW; C7EA66FBD19A1619 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNPPHGMLDLEETIAANSKD 20  
DB 86 KVNPPHGMLDLEETIAANSKD 105

## RESULT 13

ID Q9ENN7 PRELIMINARY; PRT; 251 AA.  
AC Q9ENN7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC SEQUENCE FROM N.A.  
RA STRAIN-A712;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047676; BAB12184.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28785 MW; BFE79D3A26134F18 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPPHGMLDLEETIAANSKD 20  
DB 86 KVNPPHGMLDLEETIAANSKD 105

## RESULT 14

ID Q9ENN6 PRELIMINARY; PRT; 251 AA.  
AC Q9ENN6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN-A713;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047677; BAB12185.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C34248282F19CA CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNPPHGMLDLEETIAANSKD 20  
DB 86 KVNPPHGMLDLEETIAANSKD 105

```
RESULT 15
Q9ENN5 PRELIMINARY; PRT; 251 AA.
AC Q9ENN5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A75;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047678; BAB12186.1; -.
FT NON_TER 1 1
SQ SEQUENCE 251 AA; 28658 MW; 98E8ED00B2EDF10 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVFPHGMLDEETANSKD 20
Db 86 KVFPHGMLDEETANSKD 105

RESULT 16
Q9ENN4 PRELIMINARY; PRT; 251 AA.
AC Q9ENN4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A77;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047679; BAB12187.1; -.
FT NON_TER 1 1
SQ SEQUENCE 251 AA; 28734 MW; 3B96B782882F19D9 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVFPHGMLDEETANSKD 20
Db 86 KVFPHGMLDEETANSKD 105

RESULT 17
Q9ENN2 PRELIMINARY; PRT; 251 AA.
AC Q9ENN2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
```

```
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A9;
RA Fujiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047681; BAB12189.1; -.
FT NON_TER 1 1
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 104; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVFPHGMLDEETANSKD 20
Db 86 KVFPHGMLDEETANSKD 105

RESULT 18
Q84780 PRELIMINARY; PRT; 1124 AA.
AC Q84780;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)
DE (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,
RA Rukhlina T.O., Chizhikov V.E., Petrov N.A., Prikhod'ko G.G.,
RA Blinov V.M., Vasilchenko S.K., Sandakchiev L.S., Kusov Y.Y.,
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;
RL Dokl. Biochem. 285:379-383(1986).
DR EMBL; X04200; CAA27797.1; -.
DR EMBL; A11312; CAA00953.1; -.
KW Nonstructural protein.
FT NON_TER 1 1
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 104; DB 12; Length 1124;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVFPHGMLDEETANSKD 20
Db 879 KVFPHGMLDEETANSKD 898

RESULT 19
Q05794 PRELIMINARY; PRT; 1161 AA.
AC Q05794; Q67800; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)) (FRAGMENT).
```

OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sverdlov S.D., Tsarev S.A., Marjova S.V., Vasilenko S.K.,  
RA Chizhikov V.E., Petrov N.A., Kuzov Y.Y., Nastashenko T.A.,  
RA Balayan M.S.;  
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC EMBL: X15464; CAA33492.1; -  
DR InterPro: IPR000408; RCF1.  
DR PROSITE: PS00626; RCF1\_2; UNKNOWN.1.  
KM Polypeptide: Coat protein; Core protein; RNA-directed RNA polymerase;  
KW Hydroxylase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 >1161  
FT NON\_TER 1161 1161  
SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 7, 2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNPHGMLELEIAANSKD 20  
DB 916 KVNPHGMLELEIAANSKD 935

RESULT 20  
O9WMA2 PRELIMINARY; PRT; 2216 AA.  
AC O9WMA2;  
DT 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE POLYPEPTIDE.  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AH3;  
RX MEDLINE-21386014; PubMed-11495038;  
RA Fujiwara K., Yokosuka O., Fukui K., Imazeki F., Saisho H., Omata M.,  
RA Analysis of full-length hepatitis A virus genome in sera from  
RA patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020566; BAA35104.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR004004; Calici\_pol\_hel.  
DR InterPro: IPR000408; RCF1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol.1.  
DR Pfam: PF00910; RNA\_helicase.1.  
DR PRINTS: PR00918; CALICIVIRUS.  
DR PROSITE: PS00626; RCF1\_2; UNKNOWN.1.  
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2216;  
Best Local Similarity 100.0%; Pred. No. 1, 5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KVNPHGMLELEIAANSKD 20  
DB 922 KVNPHGMLELEIAANSKD 941

RESULT 21  
O67824 PRELIMINARY; PRT; 2218 AA.  
AC O67824;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE GBM/PRNK RNA.  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GBM;  
RX MEDLINE-94076453; PubMed-8254770;  
RA Graef J., Norman A., Feinstein S.M., Flehmig B.;  
RA "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
RA to two cell culture adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75214; CAA53024.1; -  
DR InterPro: IPR000408; RCF1.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol.1.  
DR Pfam: PF00910; RNA\_helicase.1.  
DR PROSITE: PS00626; RCF1\_2; UNKNOWN.1.  
FT CHAIN 1 785  
FT CHAIN 792 1422  
FT CHAIN 1417 2218  
FT CHAIN 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;  
SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 1, 5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KVNPHGMLELEIAANSKD 20  
DB 916 KVNPHGMLELEIAANSKD 935

RESULT 22  
O9DL32 PRELIMINARY; PRT; 2225 AA.  
AC O9DL32;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE POLYPEPTIDE.  
OS Hepatitis A virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-L-A-1;  
RX Wang P.F., Jiang C.L., Liu J.Y., Zhang H.Y.;  
RX Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF314208; AAG45423.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR004004; Calici\_pol\_hel.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol.1.  
DR Pfam: PF00910; RNA\_helicase.1.

DR PRINTS: PR00918; CALICIVIRUSNS.  
SQ SEQUENCE 2225 AA; 251297 MW; EBACEA1B043E5E9B CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2225;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDEETIAANSKD 20  
|||||  
DB 922 KVNFPHGMLDEETIAANSKD 941

RESULT 23  
Q67825 PRELIMINARY; PRT: 2227 AA.  
ID 067825  
AC 067825;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GBM/WT RNA.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GBM;  
RX MEDLINE=94076453; PubMed=8254770;  
RA Graff J., Normann A., Feinstein S.M., Fleming B.;  
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
to two cell culture adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75215; CA53025.1; -  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.  
FT CHAIN 1423 2227 P2 NONSTRUCTURAL PROTEIN.  
FT CHAIN 2227 251563 P3 NONSTRUCTURAL PROTEIN.  
SQ SEQUENCE 2227 AA; 251563 MW; 4C4D79D352F936B4 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDEETIAANSKD 20  
|||||  
DB 922 KVNFPHGMLDEETIAANSKD 941

RESULT 24  
Q67826 PRELIMINARY; PRT: 2227 AA.  
ID 067826  
AC 067826;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GBM/HFS RNA.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GBM;  
RX MEDLINE=94076453; PubMed=8254770;  
RA Graff J., Normann A., Feinstein S.M., Fleming B.;  
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison

RT to two cell culture adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75216; CA53026.1; -  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.  
FT CHAIN 1423 2227 P2 NONSTRUCTURAL PROTEIN.  
FT CHAIN 2227 488CB7C962319457 P3 NONSTRUCTURAL PROTEIN.  
SQ SEQUENCE 2227 AA; 251496 MW; 488CB7C962319457 CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDEETIAANSKD 20  
|||||  
DB 922 KVNFPHGMLDEETIAANSKD 941

RESULT 25  
Q9WMA4 PRELIMINARY; PRT: 2227 AA.  
ID 09WMA4  
AC 09WMA4;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AHL;  
RX MEDLINE=21386014; PubMed=11495028;  
RA Fujiwara K., Yokosuka O., Fukui K., Imazeki F., Saisho H., Omata M.;  
RT "Analysis of full-length hepatitis A virus genome in sera from  
patients with fulminant and self-limited acute type A hepatitis.";  
RL J. Hepatol. 35:112-119(2001).  
DR EMBL: AB020564; BAA35102.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR004004; calic1\_pol\_hel.  
DR InterPro: IPR000408; RCCL.  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
DR PRINTS: PR00918; CALICIVIRUSNS.  
DR PROSITE: PS00626; RCCL\_2; UNKNOWN\_1.  
SQ SEQUENCE 2227 AA; 251304 MW; 0DEF6D2AEC29C0CE CRC64;

Query Match 100.0%; Score 104; DB 12; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDEETIAANSKD 20  
|||||  
DB 922 KVNFPHGMLDEETIAANSKD 941

Search completed: June 16, 2002, 00:08:54  
Job time: 792 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:01:47 ; Search time 209.1 Seconds  
(without alignments)  
10.624 Million cell updates/sec

Title: US-09-171-432a-45  
Perfect score: 101  
Sequence: 1 DLEETAAANSKDPFNMSETDL 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

1: A\_Geneseq\_032802.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	20	AAW42928	Immunogenic Hepat
2	101	100.0	21	AAW42928	Synthetic HAV P2A
3	101	100.0	2227	7 AAP60066	Sequence of viral
4	101	100.0	2227	11 AAR05697	Attenuated heptit
5	101	100.0	2227	18 AAW34074	Hepatitis A virus
6	101	100.0	2227	21 AAB18607	Amino acid sequenc
7	101	100.0	2227	21 AAB18608	Amino acid sequenc
8	101	100.0	2227	21 AAB18609	Amino acid sequenc
9	53	52.5	20	AAW42927	Immunogenic Hepat
10	53	52.5	20	AAW42927	Synthetic HAV P2A
11	46	45.5	424	ABW15556	Drosophila melanog

12	44.5	44.1	1712	22	ABW60536
13	43	42.6	152	18	AAW55492
14	43	42.6	157	18	AAW55327
15	43	42.6	350	20	AAW27183
16	43	42.6	536	21	AAW48090
17	43	42.6	1201	20	AAW90345
18	43	42.6	1201	20	AAW90345
19	42	41.6	89	22	ABW58421
20	42	41.6	177	21	AAW48507
21	42	41.6	226	21	AAW48506
22	42	41.6	365	18	AAW48505
23	42	41.6	365	18	AAW13493
24	42	41.6	365	20	AAW29461
25	42	41.6	402	22	ABW60260
26	42	41.6	445	22	AAW40763
27	42	41.6	445	22	AAW40764
28	42	41.6	445	22	AAW40765
29	42	41.6	538	21	AAW81713
30	42	41.6	641	22	AAW38978
31	42	41.6	698	22	AAW38977
32	42	41.6	698	22	AAW39032
33	42	41.6	705	22	AAW94135
34	42	41.6	706	22	AAW38979
35	42	41.6	1837	21	AAW85564
36	41	40.6	72	22	ABW43574
37	41	40.6	72	22	ABW26533
38	41	40.6	72	22	AAW64511
39	41	40.6	72	22	AAW77319
40	41	40.6	72	22	AAW21245
41	41	40.6	72	22	AAW37471
42	41	40.6	116	22	AAU02921
43	41	40.6	141	22	ABG17027
44	41	40.6	163	21	AAW33763
45	41	40.6	166	21	AAW33762
46	41	40.6	177	21	AAW33761
47	41	40.6	178	21	AAW54807
48	41	40.6	188	22	AAU02923
49	41	40.6	289	21	AAW16959
50	41	40.6	289	21	AAW28545
51	41	40.6	292	21	AAW16958
52	41	40.6	292	21	AAW28544
53	41	40.6	372	22	AAU02963
54	41	40.6	376	22	AAW59176
55	41	40.6	376	22	AAW11125
56	41	40.6	389	22	AAU02973
57	41	40.6	400	22	AAU02962
58	41	40.6	459	21	AAW58662
59	41	40.6	476	22	ABW65527
60	41	40.6	528	21	AAW78933

#### ALIGNMENTS

RESULT 1
AAW42928
ID AAW42928 standard; peptide: 20 AA.
XX
AC AAW42928;
XX
DT 28-APR-1998 (first entry)
XX
DE Immunogenic Hepatitis A virus peptide YK-1328.
XX
KW Immunogenic peptide: immunogenic epitope: P2A protein;
KW immune response; antibody.
XX
OS Synthetic.
OS Hepatitis A virus.
OS
PN W09740147-A1.
XX
PD 30-OCT-1997.

Drosophila melanog  
H. pylori ORF hp6e  
H. pylori ORF hp2p  
(5)-3'-hydroxy-N-m  
Arabidopsis thalia  
Drosophila sp. Cos  
Drosophila melanog  
Propionibacterium  
Arabidopsis thalia  
Arabidopsis thalia  
Penterythritol te  
Enterobacter cloac  
Drosophila melanog  
Human polypeptide  
Human polypeptide  
Human polypeptide  
Streptococcus pneu  
Human polypeptide  
Human polypeptide  
Human polypeptide  
Human protein sequ  
Human polypeptide  
Human homologue of  
Peptide #11080 enc  
Protein #8532 enc  
Human brain expres  
Human bone marrow  
Peptide #7679 enc  
Peptide #11508 enc  
Angiotensin conver  
Novel human diagno  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Thrombin inhibitor  
Human thrombin inh  
Angiotensin conver  
Angiotensin conver  
Human prostate can  
Drosophila melanog  
Human UDP-glucuron

XX 18-APR-1997; 97WO-US06891.  
XX  
XX 19-APR-1996; 96US-0015644.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fields HA, Khudyakov YE;  
XX  
XX WPI; 1997-535831/49.  
XX  
XX Immuno-genic Hepatitis A Virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX  
XX Claim 18; Page 112; 140pp; English.  
XX  
XX Peptides AAM42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 931-950, and has a reactivity of 12.5% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal.  
XX  
XX Sequence 20 AA;  
SQ

Query Match 100.0%; Score 101; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLEETIANSKDFPMSETDL 20  
DB 1 dleetiainskdfpmsetdl 20

RESULT 2  
AAB69445  
ID AAB69445 standard; Peptide; 21 AA.  
XX  
XX AAB69445;  
XX  
XX 20-APR-2001 (first entry)  
XX  
XX Synthetic HAV P2A peptide, SEQ ID NO: 45.  
DE  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
XX Hepatitis A virus.  
OS Synthetic.  
XX  
XX WO200105824-A2.  
PN  
XX 25-JAN-2001.  
PD  
XX 14-JUL-2000; 2000MO-US19267.  
PF  
XX 15-JUL-1999; 99US-0144412.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
XX Fields HA, Khudyakov YE;  
XX  
XX WPI; 2001-112681/12.  
DR  
XX  
XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines -  
XX

PS Claim 13; Page 97; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IgM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy  
XX end of the peptides enhances the IgM antibody reactivity.  
XX  
XX Sequence 21 AA;  
SQ

Query Match 100.0%; Score 101; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLEETIANSKDFPMSETDL 20  
DB 1 dleetiainskdfpmsetdl 20

RESULT 3  
AAP60066  
ID AAP60066 standard; Protein; 2227 AA.  
XX  
XX AAP60066;  
XX  
XX 26-JUN-1991 (first entry)  
XX  
XX  
DE Sequence of viral L434 polypeptide encoded by the complete  
DE nucleotide sequence of the HAV genome.  
XX  
XX Diagnosis; vaccine; passive immunotherapy.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX Key  
FH Location/Qualifiers  
FT 1..245  
FT Region /label= P1.1A  
FT 246..491  
FT Region /label= 1B  
FT 492..836  
FT Region /label= 1C  
FT 837..980  
FT Region /label= P2.2A  
FT 981..1076  
FT Region /label= 2B  
FT 1077..1422  
FT Region /label= 2C  
FT 1423..1484  
FT Region /label= P3.3A  
FT 1485..1507  
FT Region /label= 3B  
FT 1508..1678  
FT Region /label= 3C  
FT 1679..2227  
FT Region /label= 3D  
XX  
XX  
XX EP199480-A.  
PN  
XX  
XX 29-OCT-1986.  
PD  
XX 03-APR-1986; 86EP-0302465.  
PF

```

XX PR 03-APR-1985; 85US-0719329.
XX PA (CHIR-) CHIRON CORP.
XX PI Dina D, Potter SJ, Vannest GA, Caput D;
XX DR WPI: 1986-286213/44.
XX DR N-PSDB; AAN60080.
XX PT Hepatitis A virus nucleotide sequence and polypeptide - and use
XX PT in prodn. of vaccines and diagnostic probes
XX PS Claim 5; Fig 1; 18pp; English.
XX CC AAN60080 and oligonucleotide fragments are useful in detection of
XX CC hepatitis A virus; transformed hosts may be used for expression of
XX CC polypeptides and fragments useful in vaccines without risk of
XX CC infection by the virus or in prodn. of particles which are capable
XX CC of inducing immunocompetent B cells for passive immunotherapy. Pref.
XX CC epitope is derived from AAs 445-657 or 792-848 of the HAV
XX CC polypeptide sequence (AAN60066).
SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 7; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20
   ||||||||||||||||
DB 931 dleeiaanskdfpmsetcl 950

RESULT 4
AAR05697
ID AAR05697 standard; protein: 2227 AA.
XX AC AAR05697;
XX DT 15-AUG-1990 (first entry)
XX DE Attenuated hepatitis A virus.
XX KM Hepatitis A virus; vaccine; attenuated.
XX OS Hepatitis A virus, strain HM-175.
XX FH Key
XX FH Location/Qualifiers
FT 1..23
FT /label-VP4 = 1A
FT /label-VP2 = 1B
FT /label-VP3 = 1C
FT /label-VP1 = 1D
FT 792..980
FT /label-2A
FT 981..1087
FT /label-2B
FT 1088..1422
FT /label-2C
FT 1423..1496
FT /label-3A
FT 1497..1519
FT /label-3B = VPg
FT 1520..1738
FT /label-3C
FT 1739..2227
FT /label-3D
XX
```

```

PN US4894228-A.
XX PD 16-JAN-1990.
XX PF 12-JUL-1988; 88US-0217824.
XX PR 12-JUL-1988; 88US-0217824.
XX PR 12-JUL-1988; 88US-0652967.
XX PA (USSH ) US DEPT HEALTH & HUMAN.
XX PI Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;
XX PI Daemer RJ, Gust ID;
XX DR WPI: 1990-075557/10.
XX DR N-PSDB; AAO03512.
XX PT Vaccine against hepatitis A virus infection - comprises novel
XX PT attenuated hepatitis A virus strain.
XX PS Claim 1; Fig 1; 18pp; English.
XX CC The attenuated HAV is useful for inducing protective immunity against
XX CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by
XX CC several nucleotide changes distributed throughout the genome, is
XX CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is
XX CC suitable for use as an HAV vaccine. It is noted that not all the changes
XX CC are necessary for attenuation and use as a vaccine.
SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 11; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20
   ||||||||||||||||
DB 931 dleeiaanskdfpmsetcl 950

RESULT 5
AAW34074
ID AAW34074 standard; protein: 2227 AA.
XX AC AAW34074;
XX DT 27-APR-1998 (first entry)
XX DE Hepatitis A virus HM-175 protein sequence.
XX KM HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;
XX KM infection; vaccine.
XX OS Hepatitis A virus HM-175.
XX FH Key
XX FH Location/Qualifiers
FT 1..23
FT /label- VP4
FT 24..245
FT /label- VP2
FT 246..491
FT /label- VP3
FT 492..791
FT /label- VP1
FT 792..980
FT /label- 2A
FT 981..1087
FT /label- 2B
FT 1088..1422
FT /label- 2C
FT 1423..1496
FT /label- 3A
XX
```

FT Protein 1497..1519  
 FT /label= 3B  
 FT Protein 1520..1738  
 FT /label= 3C  
 FT Protein 1739..2227  
 FT /label= 3D  
 XX  
 PN MO9740166-A2.  
 XX  
 PD 30-OCT-1997.  
 XX  
 PF 18-APR-1997; 97WO-US06506.  
 XX  
 PR 19-APR-1996; 96US-0015642.  
 XX  
 PA (USSH ) US SEC DEPT HEALTH.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Emerson SU, Purcell RH, Raychaudhuri G;  
 XX  
 DR WPI; 1997-535850/49.  
 DR N-PSDB; AAT93023.  
 XX  
 PT Human attenuated HAV genome containing simian HAV 2C gene - useful  
 PT as vaccines against HAV infection  
 PS Disclosure; Fig 13A-D; 66pp; English.  
 XX  
 CC This protein sequence is encoded by the human hepatitis A virus  
 CC (HAV) HM-175 wild-type genome (see AAT93023). Attenuated strain  
 CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
 CC kidney cells. A claimed DNA construct (1) comprises a genome of  
 CC HAV, where the genome is a human attenuated HAV genome in which a  
 CC region of the 2C gene has been replaced by a corresponding region  
 CC from a 2C gene of a simian AGM-27 HAV genome (see AAT93024). The  
 CC region of the 2C gene from AGM-27 contained in the construct  
 CC preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
 CC transcript of (1); (2) a cell transfected with (1) or the RNA  
 CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
 CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
 CC its RNA transcript, can be used as a vaccine for preventing HAV in  
 CC a mammal. (1) or the RNA transcript can also be used to stimulate  
 CC the production of protective antibodies in the mammal.  
 XX  
 SQ Sequence 2227 AA:  
 XX  
 Query Match 100.0%; Score 101; DB 18; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEIRANSKDFPMSETDL 20  
 ||||||||||||||||  
 DB 931 dleeianskdfpmsetdl 950  
 XX  
 RESULT 6  
 AAB18607  
 ID AAB18607 standard; Protein; 2227 AA.  
 XX  
 AC AAB18607;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 XX  
 KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX  
 PT

PD 05-SEP-2000.  
 XX  
 PF 07-JUN-1995; 95US-0475886.  
 XX  
 PR 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93WO-US08610.  
 PR 10-MAR-1995; 95US-0397232.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX  
 DR WPI; 2000-586464/55.  
 DR N-PSDB; AAA75476.  
 XX  
 PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type  
 XX  
 PS Disclosure; Fig 6A-K; 72pp; English.  
 XX  
 CC The present sequence is derived from a wild type hepatitis A virus  
 CC (HAV) strain HM-174. The sequence is modified to produce HAV which  
 CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
 CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
 CC attenuation. It is useful as a live vaccine for prophylaxis of  
 CC hepatitis A in humans and other primates.  
 XX  
 SQ Sequence 2227 AA:  
 XX  
 Query Match 100.0%; Score 101; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEIRANSKDFPMSETDL 20  
 ||||||||||||||||  
 DB 931 dleeianskdfpmsetdl 950  
 XX  
 RESULT 7  
 AAB18608  
 ID AAB18608 standard; Protein; 2227 AA.  
 XX  
 AC AAB18608;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
 XX  
 KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KM P-35 virus.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX  
 PD 05-SEP-2000.  
 PD 07-JUN-1995; 95US-0475886.  
 XX  
 PR 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93WO-US08610.  
 PR 10-MAR-1995; 95US-0397232.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX  
 DR WPI; 2000-586464/55.  
 DR N-PSDB; AAA75477.  
 XX  
 PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type  
XX  
PS Disclosure; Columns 67-78; 72pp; English.  
XX  
CC The present sequence is derived from passage 35 of a wild type  
CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA:  
  
Query Match 100.0%; Score 101; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLEETIAANSKDFPNMSEITDL 20  
Db 931 dleeaanskfpmsetdl 950  
|||||  
RESULT 8  
AAB18609  
ID AAB18609 standard; Protein; 2227 AA.  
XX  
AC AAB18609;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of live attenuated Hepatitis A virus 4380.  
XX  
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
KM HAV 4380.  
XX  
OS Hepatitis A virus.  
XX  
PN US6113912-A.  
XX  
PD 05-SEP-2000.  
XX  
PF 07-JUN-1995; 95US-0475886.  
XX  
PR 18-SEP-1992; 92US-0947338.  
PR 17-SEP-1993; 93WO-US08610.  
PR 10-MAR-1993; 95US-0397232.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX  
DR WPI; 2000-586464/55.  
DR N-PSDB; AAA75478.  
XX  
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type  
XX  
PS Disclosure; Columns 93-104; 72pp; English.  
XX  
CC The present sequence is derived from a live attenuated hepatitis A  
CC virus (HAV) of the invention, designated HAV 4380. The sequence is  
CC produced by modifying wild type HAV strain HM-174. The HAV of the  
CC invention are adapted to growth in the human fibroblast-like cell  
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain  
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis  
CC of hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA:

Query Match 100.0%; Score 101; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLEETIAANSKDFPNMSEITDL 20  
Db 931 dleeaanskfpmsetdl 950  
|||||  
RESULT 9  
AAW42927  
ID AAW42927 standard; Peptide; 20 AA.  
XX  
AC AAW42927;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1327.  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1327.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
FN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudiyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
CC Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 922-941, and has a reactivity of 31.3% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal.  
XX  
SQ Sequence 20 AA:  
  
Query Match 52.5%; Score 53; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.063;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLEETIAANSKD 11  
Db 10 dleeaanskd 20  
|||||  
RESULT 10  
AAB69444  
ID AAB69444 standard; Peptide; 21 AA.  
XX

AC	AAB69444;
XX	
DF	20-APR-2001 (first entry)
XX	
DE	Synthetic HAV p2A peptide, SEQ ID NO: 44.
XX	
KW	Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;
KM	antigen; major structural capsid polypeptide; HAV antibody detection.
XX	
OS	Hepatitis A virus.
OS	Synthetic.
FN	WO200105824-A2.
PD	
XX	25-JAN-2001.
XX	
FP	14-JUL-2000; 2000MO-USI9267.
XX	
PR	15-JUL-1999; 99US-0144412.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	
P1	Fields HA, Khudjakov YE;
XX	
DR	WPI; 2001-112681/12.
XX	
PT	Synthetic peptides used as antigen sources for enzyme immunoassays
PF	detecting anti-hepatitis A virus and as vaccines -
XX	
PS	Claim 13; Page 96; 130pp; English.
XX	
CC	The present sequence is one of a number of synthetic peptides which are
CC	immunoreactive with hepatitis A virus (HAV) antibodies. The peptides
CC	comprise antigenic epitopes of the major structural capsid polypeptides
CC	or non-structural polypeptides of HAV with one or more glutamine
CC	molecules at the carboxy end of the peptide. The peptides are used to
CC	detect the presence of antibodies against HAV in mammalian serum, to
CC	detect the presence of HAV in a human or animal through the binding of
CC	the peptide to an antibody, to detect acute phase infection by detecting
CC	Igm antibodies in mammalian serum and detecting convalescence in a
CC	mammal. The peptides are used to detect or quantify HAV antibodies in
CC	samples in clinical or research-based assays using immunoblotting,
CC	fluorescence in situ hybridisation analysis, gel-mobility shift assays,
CC	tracking of radioactive or bioluminescent markers, chromatography or
CC	electrophoresis. The peptides are used to induce an immune response to
CC	HAV when administered to a human or animal. Glutamine at the carboxy
CC	end of the peptides enhances the Igm antibody reactivity.
XX	
SO	Sequence 21 AA:
XX	
OY	1 DLEETAAASKD 11
DB	10 dleetaanskd 20
XX	
RESULT 11	
ABB71556	
ID	ABB71556 standard; Protein; 424 AA.
AC	
XX	ABB71556;
DT	
XX	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 41460.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KM	pharmaceutical.
XX	

OS	Drosophila melanogaster.
XX	
FN	MO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PE	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PA	11-JUL-2000; 2000US-0614150.
XX	
(PEKE )	PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
DR	WPI; 2001-656860/75.
DR	N-PSDB; ABL15659.
PT	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
PS	Disclosure: SEQ ID NO 41460; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABB57737-ABB72072).
CC	
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	Sequence 424 AA:
SQ	
Query Match	45.5%; Score 46; DB 22; Length 424;
Best Local Similarity	47.1%; Pred. NO. 30;
Matches	8; Conservative 6; Mismatches 3; Indels 0; Gaps 0
OY	1 DIEETANSKDFPNMSE 17 I : I : I I I : I : I :
Db	180 dvaelaamsrdmphiinak 196
RESULT 12	
ABB60536	
ID	ABB60536 standard; Protein; 1712 AA.
XX	
AC	ABB60536;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 8400.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
FN	MO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
ER	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
PA	(PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI: 2001-656860/75.  
DR N-PSDB: ABL04639.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 8400; 21np + Sequence listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157737-AB172072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1712 AA;  
  
Query Match 44.1%; Score 44.5; DB 22; Length 1712;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;  
  
OY 1 DLEE-IAANSKDFPMSETD 19  
Db 44 diqliaetgqdlpusedt 63  
::: || | || || ||  
|::: || | || || ||  
  
RESULT 13  
AAW53492  
ID AAW55492 standard; Protein: 152 AA.  
XX  
AC AAW55492;  
XX  
DT 30-JUN-1998 (first entry)  
XX  
XX H. pylori ORF hp6e10967\_24889750\_f2\_7 secreted protein.  
XX  
DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW Identification; binding compound; bacteria; life cycle; activator;  
KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
KW bacterium.  
XX  
XX Helicobacter pylori.  
XX  
OS  
XX W09737044-A1.  
XX  
PD 09-OCT-1997.  
XX  
PE 27-MAR-1997; 97WO-US05223.  
XX  
PF 06-DEC-1996; 96US-0761318.  
PR 29-MAR-1996; 96US-0625811.  
PR 02-APR-1996; 96US-0758731.  
PR 25-OCT-1996; 96US-0736905.  
PR 28-OCT-1996; 96US-0738859.  
XX  
PA (ASTR ) ASTRA AB.  
XX  
PI Alm RA, Smith D;  
XX  
DR WPI: 1997-503122/46.  
DR N-PSDB: AAV24901.  
XX  
PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection  
XX  
PS Claims 14,94; Page 698; 1145pp; English.  
XX  
CC This sequence is a H. pylori secreted protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The DNA and probes derived from it may be used for the  
CC identification of H. pylori in a sample, and the diagnosis of  
CC H. pylori infection. Nucleic acid sequences complementary to the  
CC DNA act as antisense sequences, and can be used to prevent the  
CC translation of H. pylori mRNA. Antibodies against the protein can  
CC be used in immunoassays to evaluate the abundance and distribution  
CC of H. pylori-specific antigens. The genomic sequence of H. pylori  
CC (ATCC 55679) was determined from overlapping contigs generated  
CC by mechanically shearing the bacterial DNA. The sequences were  
CC analysed for ORF of at least 180 nucleotides, and the predicted  
CC coding regions defined by computer evaluation. To identify likely  
CC H. pylori antigens for vaccine development, the amino acid  
CC sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having  
CC identified and determined the sequences of interest, particular  
CC regions can be isolated from H. pylori by PCR amplification for  
CC recombinant polypeptide production, e.g. in *E. coli* hosts.  
XX  
SQ Sequence 152 AA;  
  
Query Match 42.6%; Score 43; DB 18; Length 152;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
OY 5 IAANSKDFPMSETDL 20  
Db 16 valnakdtskdsedl 31  
: || || | : ||  
|::: || | : ||  
  
RESULT 14  
AAW53327  
ID AAW55327 standard; Protein: 157 AA.  
XX  
AC AAW53327;  
XX  
DT 15-JUN-1998 (first entry)  
XX  
XX H. pylori ORF hp2p10625orf6 protein.  
XX  
DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW Identification; binding compound; bacteria; life cycle; activator;  
KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
XX  
XX Helicobacter pylori.  
XX  
OS  
XX W09737044-A1.  
XX  
PD 09-OCT-1997.  
XX  
PE 27-MAR-1997; 97WO-US05223.  
XX  
PF 06-DEC-1996; 96US-0761318.  
PR 29-MAR-1996; 96US-0625811.  
PR 02-APR-1996; 96US-0758731.  
PR 25-OCT-1996; 96US-0736905.  
PR 28-OCT-1996; 96US-0738859.  
XX  
PA (ASTR ) ASTRA AB.  
XX  
PI Alm RA, Smith D;  
XX  
DR WPI: 1997-503122/46.  
DR N-PSDB: AAV24736.





PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140693.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142153.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148310.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149420.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160860.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.



CC in producing compositions that modulate the expression or function of the  
CC encoded protein, for gene therapy, mapping functional regions of the  
CC protein and in studying associated physiological pathways. In addition,  
CC modulation of *Cos2* gene activity is used for prophylactic and therapeutic  
CC purposes such as treatment of cancers (e.g. basal cell carcinoma, glioma,  
CC medulloblastoma, meningioma, breast carcinoma), hedgehog signalling  
CC pathway function, identification of cell type based on expression,  
CC infertility treatment, contraception, developmental abnormalities. The  
CC encoded protein is useful as an immunogen for producing specific  
CC antibodies, in screening for biologically active agents that modulate  
CC *Cos2* function and that act in the hedgehog signalling pathway. The  
CC protein can also be used for therapeutic and prophylactic purposes.  
XX  
SQ Sequence 1201 AA;  
  
Query Match 42.6%; Score 43; DB 20; Length 1201;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
OY 4 EIAANSKDFPMNSEPD 19  
:: | | | | | | | | | |  
Db 48 qvpaekdnpisecd 63  
  
RESULT 18  
ABBS8421  
ID ABB58421 standard; Protein: 1201 AA.  
XX  
AC ABB58421:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 2055.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB: ABL02524.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 2055; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX SQ Sequence 1201 AA;  
  
Query Match 42.6%; Score 43; DB 22; Length 1201;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
OY 4 EIAANSKDFPMNSEPD 19  
:: | | | | | | | | | |  
Db 48 qvpaekdnpisecd 63  
  
RESULT 19  
AAU59799  
ID AAU59799 standard; Protein: 89 AA.  
XX  
AC AAU59799:  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #20695.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
XX 02-JUN-2000; 2000US-208841P.  
XX 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIAX CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI: 2001-616774/71.  
DR N-PSDB: AAS59606.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID NO 20994; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pst\_sequences.  
XX Sequence 89 AA:

Query Match 41.6%; Score 42; DB 22; Length 89;  
Best Local Similarity 42.1%; Prd. No. 23;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 DLEETIANSKDPNNSETD 19  
|||:|:|:|:|  
Db 31 dlemwagrgvgnfnlfgvd 49

RESULT 20

AAC48507  
ID AAC48507 standard; Protein; 177 AA.

XX AAC48507;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 61261.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0149932.  
PR 26-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156536.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 42; DB 21; Length 177;  
Best Local Similarity 36.8%; Pred. No. 50;

Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
Oy 1 DLEIAANSKDPNMSETD 19  
||: ||: :| :| :|  
Db 157 diaplanahsygmhvd 175  
  
RESULT 21  
AAG48506  
ID AAG48506 standard; Protein: 226 AA.  
XX  
AAG48506;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SPQ ID NO: 61260.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
PN  
XX  
XX  
PD 06-SEP-2000.  
PE  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130444.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136783.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143297.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149729.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151086.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161820.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 42; DB 21; Length 226;  
Best Local Similarity 36.8%; Pred. No. 66;  
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
OY 1 DLIEIANSKDPNMSFTD 19  
Db 206 dlapvianahsypgmhvd 224

```
RESULT 22
AAG48505
ID AAG48505 standard; Protein: 265 AA.
AC
AC AAG48505:
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61259.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
FD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
```

```
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 42; DB 21; Length 265;
Best Local Similarity 36.8%; Pred. No. 79;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 DLEETANSKDFPMNSETD 19
DB 245 dlaplanahsygmnmvd 263

RESULT 23
AAW13493
ID AAW13493 standard; Protein; 365 AA.
AC AAW13493;
XX
XX
XX 09-JUL-1997 (first entry)
DT
```

```
XX DE Pentaerythritol tetranitrate reductase enzyme.
XX KM PENTN: E. cloacae; explosive; bioremediation;
XX KM glycerol trinitrate; GTN; ethylene glycol dinitrate; EGDN.
XX OS Enterobacter cloacae strain PB2.
XX PN WO9703201-A1.
XX PD 30-JAN-1997.
XX PE 08-JUL-1996; 96WO-GB01629.
XX PR 04-APR-1996; 96GB-0007220.
XX PR 11-JUL-1995; 95GB-0014138.
XX PA (MINA ) UK SEC FOR DEFENCE.
XX PI Blinks PR, Bruce NC, French CE, Nicklin S;
XX DR WPI: 1997-119056/11.
XX DR N-PSDB; AAT61540.
XX PS Claim 3; Fig 4; 39pp; English.
XX CC The present sequence represents novel protein pentaerythritol
XX CC tetranitrate (PETN) reductase enzyme. This enzyme catalyses the removal
XX CC of nitrate from PETN, and has reductase activity specifically at the
XX CC nitrate ester linkage of PETN. The PETN reductase enzyme can be used for
XX CC bioremedial treatment of an environment, especially a waste stream, soil
XX CC or ground sample, contaminated by PETN, glycerol trinitrate (GTN) and/or
XX CC ethylene glycol dinitrate (EGDN). The enzyme degrades the contaminant
XX CC which allows the PETN, GTN and/or EGDN present in the material to be
XX CC consumed. The enzyme can also be used for producing pentaerythritol
XX CC di- and tri-nitrates by reacting PETN with the enzyme, and for the
XX CC detection of PETN, GTN and/or EGDN in a sample, preferably as a
XX CC component of a biosensor.
XX SQ Sequence 365 AA;

Query Match 41.6%; Score 42; DB 18; Length 365;
Best Local Similarity 47.4%; Pred. No. 11e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 LEEETANSKDFPMNSETDL 20
DB 258 leelakrglayllhmsedl 276

RESULT 24
AAV29461
ID AAV29461 standard; Protein; 365 AA.
XX
XX AAV29461;
XX
XX DT 05-OCT-1999 (first entry)
XX DE Enterobacter cloacae PB2 pentaerythritol tetranitrate reductase.
XX DE
XX KM Enterobacter cloacae PB2; pentaerythritol tetranitrate reductase;
XX KM PETN; degradation; explosive; trinitrotoluene; TNT; bioremediation;
XX KM environmental waste; water; soil.
XX OS Enterobacter cloacae.
XX
XX FT Key Location/Qualifiers
FT Peptide 1
```



```

FT      Protein      /label= signal
FT      2..365      /label= Pentaerythritol_tetranitrate_reductase
XX      XX
XX      GB2332432-A.
XX      PD
XX      23-JUN-1999.
XX      PF
XX      19-DEC-1997: 97GB-0026813.
XX      PR
XX      19-DEC-1997: 97GB-0026813.
XX      PA
XX      (MINA ) UK SEC FOR DEFENCE.
XX      PI
XX      Bruce NC, French CE, Nicklin S;
XX      WPI: 1999-421960/36.
XX      DR
XX      N-PSDB; AAX90586.
XX      PT
XX      Enterobacter able to degrade trinitrotoluene
XX      PS
XX      Claim 4; Page 14-15; 22pp; English.
XX      CC
XX      The present sequence represents pentaerythritol tetranitrate reductase
XX      (PERN) isolated from an enterobacter cloacae bacterial strain referred
XX      to as PB2 and deposited as NCIMB 40718, which is able to degrade
XX      trinitrotoluene (TNT). The enterobacter cloacae bacterial strain is
XX      useful for the preparation of compositions for the biodegradation of
XX      environmental waste containing TNT.
XX      SQ
XX      Sequence 365 AA:

Query Match      41.6%; Score 42; DB 20; Length 365;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      2 LBEIANSKDFPNKSETDL 20
       :|:|:| :|:|:|:|
Db      258 leelakrglaylmsetdl 276

RESULT 25
ABB60260
ID      ABB60260 standard; Protein; 402 AA.
XX
XX      ABB60260;
XX      AC
XX      26-MAR-2002 (first entry)
XX      DT
XX      Drosophila melanogaster polypeptide SEQ ID NO 7572.
XX      DE
XX      Drosophila: developmental biology; cell signalling; insecticide;
XX      KM
XX      pharmaceutical.
XX      OS
XX      Drosophila melanogaster.
XX      OS
XX      WO200171042-A2.
XX      PN
XX      27-SEP-2001.
XX      PD
XX      23-MAR-2001; 2001WO-US09231.
XX      PF
XX      23-MAR-2000; 2000US-191637P.
XX      PR
XX      11-JUL-2000; 2000US-0614150.
XX      PA
XX      (PEKE ) PE CORP NY.
XX      PI
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX      WPI: 2001-656860/75.
XX      DR
XX      N-PSDB; ABL04363.
XX

```

```

PT      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX      PS
XX      Disclosure; SEQ ID NO 7572; 21pp + Sequence Listing; English.
XX      CC
XX      The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signalling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutics and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX      sequences (ABL01840-ABL16175) and the encoded proteins
XX      (ABB57737-ABB72072).
XX      CC
XX      The sequence data for this patent did not form part of the printed
XX      CC      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX      SQ
XX      Sequence 402 AA:

Query Match      41.6%; Score 42; DB 22; Length 402;
Best Local Similarity 53.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY      2 LBEIANSKDFPN 14
       :|:|:| |:|:|:|
Db      117 leevsamsknfpd 129

Search completed: June 16, 2002, 00:01:48
Job time: 12737 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:19 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec.

Title: US-09-171-432a-45  
Perfect score: 101  
Sequence: 1 DLEIRIANSKDFPNMSETDL 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 60 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	2227	3	US-08-475-886-2	Sequence 2, Appl
2	101	100.0	2227	3	US-08-475-886-4	Sequence 4, Appl
3	101	100.0	2227	3	US-08-475-886-6	Sequence 6, Appl
4	101	100.0	2227	4	US-08-397-232-2	Sequence 2, Appl
5	101	100.0	2227	4	US-08-397-232-4	Sequence 4, Appl
6	101	100.0	2227	4	US-09-171-387-2	Sequence 2, Appl
7	44	43.6	907	3	US-08-938-830-26	Sequence 26, Appl
8	44	43.6	907	3	US-09-020-222-26	Sequence 26, Appl
9	43	42.6	1201	4	US-09-098-901-2	Sequence 26, Appl
10	42	41.6	365	2	US-08-983-352-2	Sequence 2, Appl
11	41	40.6	376	4	US-09-200-965-2	Sequence 2, Appl
12	39	38.6	50	1	US-08-127-351-14	Sequence 14, Appl
13	39	38.6	50	1	US-08-480-367B-14	Sequence 14, Appl
14	39	38.6	50	1	US-08-487-221A-14	Sequence 14, Appl
15	39	38.6	50	1	US-08-480-370-14	Sequence 14, Appl
16	39	38.6	352	4	US-09-286-691-26	Sequence 26, Appl
17	39	38.6	352	4	US-09-687-147-26	Sequence 26, Appl
18	39	38.6	396	1	US-08-430-024-2	Sequence 2, Appl
19	39	38.6	396	1	US-08-782-009-2	Sequence 2, Appl
20	39	38.6	396	3	US-09-017-302-2	Sequence 2, Appl
21	39	38.6	984	1	US-08-242-932-2	Sequence 2, Appl
22	39	38.6	984	1	US-08-714-481-2	Sequence 2, Appl
23	39	38.6	984	1	PCT-US95-06111-2	Sequence 2, Appl
24	39	38.6	1098	4	US-08-923-992A-8	Sequence 8, Appl
25	39	38.6	1104	4	US-08-923-992A-4	Sequence 4, Appl
26	39	38.6	1128	4	US-08-923-992A-6	Sequence 6, Appl
27	39	38.6	1164	4	US-08-923-992A-2	Sequence 2, Appl

28	39	38.6	1164	4	US-08-923-992A-10	Sequence 10, Appl
29	38	37.6	99	1	US-08-004-492-9	Sequence 9, Appl
30	38	37.6	99	2	US-08-484-905-60	Sequence 60, Appl
31	38	37.6	99	3	US-08-481-985B-60	Sequence 60, Appl
32	38	37.6	99	4	US-08-370-476-60	Sequence 60, Appl
33	38	37.6	105	1	US-08-081-539-73	Sequence 73, Appl
34	38	37.6	105	1	US-08-466-647-73	Sequence 73, Appl
35	38	37.6	114	4	US-09-347-819-4	Sequence 4, Appl
36	38	37.6	162	2	US-08-716-317-10	Sequence 10, Appl
37	38	37.6	163	2	US-08-716-317-15	Sequence 15, Appl
38	38	37.6	164	1	US-08-318-193-60	Sequence 60, Appl
39	38	37.6	172	4	US-08-149-101A-21	Sequence 21, Appl
40	38	37.6	172	5	PCT-US94-12873-21	Sequence 21, Appl
41	38	37.6	174	4	US-08-149-101A-20	Sequence 20, Appl
42	38	37.6	174	5	PCT-US94-12873-20	Sequence 20, Appl
43	38	37.6	181	1	US-08-209-182C-4	Sequence 4, Appl
44	38	37.6	181	5	PCT-US92-05612-4	Sequence 4, Appl
45	38	37.6	182	4	US-08-149-101A-25	Sequence 25, Appl
46	38	37.6	182	5	PCT-US94-12873-25	Sequence 25, Appl
47	38	37.6	183	1	US-08-009-973-1	Sequence 1, Appl
48	38	37.6	184	1	US-08-567-047-2	Sequence 2, Appl
49	38	37.6	184	2	US-08-693-182-2	Sequence 2, Appl
50	38	37.6	184	2	US-08-567-048-2	Sequence 2, Appl
51	38	37.6	184	2	US-09-008-482-2	Sequence 2, Appl
52	38	37.6	184	2	US-08-945-529-8	Sequence 8, Appl
53	38	37.6	184	2	US-08-945-529-10	Sequence 10, Appl
54	38	37.6	184	2	US-08-945-529-11	Sequence 11, Appl
55	38	37.6	184	2	US-08-149-101A-17	Sequence 17, Appl
56	38	37.6	184	4	US-08-149-101A-22	Sequence 22, Appl
57	38	37.6	184	4	PCT-US94-12873-17	Sequence 17, Appl
58	38	37.6	184	5	PCT-US94-12873-22	Sequence 22, Appl
59	38	37.6	184	5	PCT-US94-12873-22	Sequence 22, Appl
60	38	37.6	184	6	5186931-1	Patent No. 5186931

## ALIGNMENTS

RESULT 1  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OR INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 202642620S2  
; CURRENT APPLICATION NUMBER: US/08/475, 886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947, 338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397, 232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2,3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLEIRIANSKDFPNMSETDL 20  
Db 931 DLEIRIANSKDFPNMSETDL 950

```

RESULT 2
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

```

```

Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 DLEETANSKDFPNMSETDL 20
|||||
Db 931 DLEETANSKDFPNMSETDL 950

```

```

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

```

```

Query Match          100.0%; Score 101; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 DLEETANSKDFPNMSETDL 20
|||||
Db 931 DLEETANSKDFPNMSETDL 950

```

```

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A

```

```

; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

```

```

Query Match          100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 DLEETANSKDFPNMSETDL 20
|||||
Db 931 DLEETANSKDFPNMSETDL 950

```

```

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; EARLIER FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

```

```

Query Match          100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 DLEETANSKDFPNMSETDL 20
|||||
Db 931 DLEETANSKDFPNMSETDL 950

```

```

RESULT 6
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHAUDHURI, GOPA;

```

EMERSON, SUZANNE, U.:  
PORCELL, ROBERT, H.  
TITLE OF INVENTION: SIMIAN-HUMAN HAV  
HAVING A CHIMERIC 2C PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,387  
FILING DATE: 24-Mar-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/06506  
FILING DATE: 18-Apr-1997  
APPLICATION NUMBER: US60/015,642  
FILING DATE: 19-Apr-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feller  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4229U51  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2227 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 2  
US-09-171-387-2

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2,3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEIANSKDFPMNSETDL 20  
Db 931 DLEIANSKDFPMNSETDL 950

RESULT 7  
US-08-938-830-26  
Sequence 26, Application US/08938830  
Patent No. 6040437  
GENERAL INFORMATION:  
APPLICANT: Lasky, Laurence A.  
APPLICANT: Dowbenko, Donald J.  
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPalcin (Genentech)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,830  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/798419  
FILING DATE: 07-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1066P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-3216  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-938-830-26

Query Match 43.6%; Score 44; DB 3; Length 907;  
Best Local Similarity 47.1%; Pred. No. 30;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LEEIANSKDFPMNSET 18  
Db 49 LQELAASSADPEVGST 65

RESULT 8  
US-09-020-222-26  
Sequence 26, Application US/09020222  
Patent No. 6111073  
GENERAL INFORMATION:  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPalcin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/020,222  
FILING DATE: 06-Feb-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/798419  
FILING DATE: 02/07/1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1066-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-3216  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-020-222-26

Query Match 43.6%; Score 44; DB 3; Length 907;  
Best Local Similarity 47.1%; Pred. No. 30;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 LEEIANSKDFPMSETD 18  
DB 49 LOELASSADIPVSGT 65

## RESULT 9

US-09-098-901-2  
Sequence 2, Application US/09098901B  
Patent No. 6218144  
GENERAL INFORMATION:  
APPLICANT: Scott, Matthew  
APPLICANT: Slsson, John C.  
TITLE OF INVENTION: Costal2 Genes and their Uses  
FILE REFERENCE: SUN-65P  
CURRENT APPLICATION NUMBER: US/09/098,901B  
CURRENT FILING DATE: 1998-06-17  
EARLIER APPLICATION NUMBER: 60/051,347  
EARLIER FILING DATE: 1997-06-30  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1201  
TYPE: PRT  
ORGANISM: D. melanogaster  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(1201)  
OTHER INFORMATION: Xaa - Any Amino Acid  
US-09-098-901-2

Query Match 42.6%; Score 43; DB 4; Length 1201;  
Best Local Similarity 50.0%; Pred. No. 63;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 4 EIAANSKDFPMSETD 19  
DB 48 QYPAEKDPMSETD 63

## RESULT 10

US-08-983-352-2  
Sequence 2, Application US/08983352  
Patent No. 5928859  
GENERAL INFORMATION:  
APPLICANT: Nicklin, Stephen  
APPLICANT: Binks, Peter R.  
APPLICANT: Bruce, Neil C.  
APPLICANT: French, Christopher E.  
TITLE OF INVENTION: DETECTION AND BIODEGRADATION OF  
TITLE OF INVENTION: EXPLOSIVES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 No. 5928859th Giebe Rd. 8th floor  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22201-4741  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-10S  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/983,352  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/01629  
FILING DATE: 08-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9514138.8  
FILING DATE: 11-JUL-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9607220.2  
FILING DATE: 04-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 124-617  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-983-352-2

Query Match 41.6%; Score 42; DB 2; Length 365;  
Best Local Similarity 47.4%; Pred. No. 22;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 LEEIANSKDFPMSETD 20  
DB 258 IEELARKGIVLHMSIDL 276

## RESULT 11

US-09-200-965-2  
Sequence 2, Application US/09200965  
Patent No. 6133422  
GENERAL INFORMATION:  
APPLICANT: ROSEN, ET AL.  
TITLE OF INVENTION: Thrombin Inhibitor  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/200,965  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/171,817  
FILING DATE: 22-DECEMBER-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-47  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 AMINO ACIDS  
TYPE: AMINO ACID

STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-200-965-2

Query Match 40.6%; Score 41; DB 4; Length 376;  
Best Local Similarity 33.3%; Pred. No. 34;  
Matches 10; Conservative 3; Mismatches 7; Indels 10; Gaps 1;

QY 1 DLEETAN-----SKFPPNSETDL 20  
|:|:| 11 11:111  
Db 281 DMESVRLNIGMTDAFELCKADPFGSMQFDL 310

RESULT 12  
US-08-127-351-14  
Sequence 14, Application US/08127351  
Patent No. 5449761  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,351  
FILING DATE: 28-SEP-1993  
CLASSIFICATION: 534  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-127-351-14

Query Match 38.6%; Score 39; DB 1; Length 50;  
Best Local Similarity 43.8%; Pred. No. 6.4;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 EIAANSKDFPNMSETD 19  
|:|:|:| 11:111  
Db 15 ELAENNLNLPKMAEKD 30

RESULT 13  
US-08-480-367B-14

Sequence 14, Application US/08480367B  
Patent No. 5578288  
GENERAL INFORMATION:

APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,367B  
FILING DATE: 07-06-95  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 2654-002A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 684-1111  
TELEFAX: (703) 684-1124  
TELEX:

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-480-367B-14

Query Match 38.6%; Score 39; DB 1; Length 50;  
Best Local Similarity 43.8%; Pred. No. 6.4;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 EIAANSKDFPNMSETD 19  
|:|:|:| 11:111  
Db 15 ELAENNLNLPKMAEKD 30

RESULT 14  
US-08-487-221A-14  
Sequence 14, Application US/08487221A  
Patent No. 5593656  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:

```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/487,221A
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE: 28-SEP-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Villacorta, Gilberto M.
? REGISTRATION NUMBER: 34,038
? REFERENCE/DOCKET NUMBER: 4980-004-44
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 413-3000
? TELEFAX: (703) 413-2220
? TELERX: 248855 OPAT UR
? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 50 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
? MOLECULE TYPE: peptide
? US-08-487-221A-14

```

Query Match	38.6%	Score 39	DB 1	Length 50
Best Local Similarity	43.8%	Pred. No. 6.4		
Matches	7	Conservative	4	Mismatches 5
				Indels 0
				Gaps 0
QY	4	ETIANSKDEPNMSFTD	19	
		: : : : : :		
DB	15	ETIENNENLNPKEAEKD	30	

RESULT 15  
US-08-480-370-14  
; Sequence 14, Application US/08480370  
; Patent No. 5609847

1 GENERAL INFORMATION:  
2 APPLICANT: BELINKA JR, BENJAMIN A.  
3 APPLICANT: COUGHLIN, DANIEL J.  
4 APPLICANT: ALVAREZ, VERNON L.  
5 APPLICANT: WOOD, RICHARD  
6 TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDES  
7 TITLE OF INVENTION: CONSTRUCTS  
8 NUMBER OF SEQUENCES: 56  
9 CORRESPONDENCE ADDRESSES:  
10 ADDRESSEE: OHION, SPIVAK, MCLELLAND, MAIER &  
11 ADDRESSEE: NEUSTADT,  
12 ADDRESSEE: P.C.  
13 STREET: 1755 S. Jefferson Davis Highway, Suite 400  
14 City: Arlington  
15 STATE: Virginia  
16 COUNTRY: U.S.A.  
17 ZIP: 22202  
18  
19 COMPUTER READABLE FORM:  
20 MEDIUM TYPE: Floppy disk  
21 COMPUTER: IBM PC compatible  
22 OPERATING SYSTEM: PC-DOS/MS-DOS  
23 SOFTWARE: PatentIn Release #1.0, Version #1.25  
24 CURRENT APPLICATION DATA:  
25 APPLICATION NUMBER: US/08/480,370  
26 FILING DATE:  
27 CLASSIFICATION: 424  
28 PRIOR APPLICATION DATA:  
29 APPLICATION NUMBER: US 08/127,351  
30 FILING DATE: 28-SEP-1993  
31 ATTORNEY/AGENT INFORMATION:  
32 NAME: Villacorta, Gilberto M.  
33 REGISTRATION NUMBER: 34,038

```

? REFERENCE/DOCKET NUMBER: 4980-004-44
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 413-3000
? TELEFAX: (703) 413-2220
? TELE: 248655 OPAT UR
? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 50 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
? MOLECULE TYPE: peptide
US-08-480-370-14

```

Query Match	38.6%	Score 39;	DB 1;	Length 50;
Best Local Similarity	43.8%;	Pred. No. 6.4;		
Matches	7;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;

```
QY      4 EIAANSKDFPNMSETD 19
          | : | | : : | | : |
Db     15 ELAENNLLPKMAEKD 30
```

RESULT 16  
US-09-286-691-26  
; Sequence 26, Application US/09286691

```

1  GENERAL INFORMATION:
2  APPLICANT: Li, Xin-Liang
3  APPLICANT: Ljungdahl, Lars G.
4  APPLICANT: Chen, Huizhong
5  TITLE OF INVENTION: Cellulases and Coding Sequences
6  FILE REFERENCE: 42-96
7  CURRENT APPLICATION NUMBER: US/09/286,691
8  CURRENT FILING DATE: 1999-04-05
9  EARLIER APPLICATION NUMBER: US 60/027,883
10 EARLIER FILING DATE: 1996-10-04
11 EARLIER APPLICATION NUMBER: PCT US97/18008
12 EARLIER FILING DATE: 1997-10-03
13 NUMBER OF SEQ ID NOS: 29
14 SOFTWARE: PatentIn Ver. 2.0
15 SEQ ID NO 26
16 LENGTH: 352
17 TYPE: PRN
18 ORGANISM: Agaricus bisporus
19 OS-09-286-691-26

```

Query Match	38.68;	Score 39;	DB 4;	Length 352;
Best Local Similarity	46.78;	Pred. No. 68;		
Matches	7;	Conservative	5;	Mismatches 3;
				Indels 0;
				Gaps 0;

```
QY 2 LEEIANSKDFPNMS 16
    ::||| | ||::|
Db 107 VDQIAQIKQFPDVS 121
```

RESULT 17  
US-09-687-147-26  
: Sequence 26, Application US/09687147  
: Patent No. 6268198  
: GENERAL INFORMATION:  
: APPLICANT: Li, Xin-liang  
: APPLICANT: Ljungdahl, Lars G.  
: APPLICANT: Chen, Huizhong  
: TITLE OF INVENTION: Cellulases and Coding Sequences  
: FILE REFERENCE: 42-96a  
: CURRENT APPLICATION NUMBER: US/09/687,147  
: CURRENT FILING DATE: 2000-10-12  
: PRIOR APPLICATION NUMBER: US 60/027,883  
: PRIOR FILING DATE: 1996-10-04  
: PRIOR APPLICATION NUMBER: PCT US97/18008  
: PRIOR FILING DATE: 1997-10-03



;; PRIOR APPLICATION NUMBER: 09/286,691  
;; PRIOR FILING DATE: 1999-04-05  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 26  
;; LENGTH: 352  
;; TYPE: PRT  
;; ORGANISM: Agaricus bisporus  
US-09-687-147-26

Query Match 38.6%; Score 39; DB 4; Length 352;  
Best Local Similarity 46.7%; Pred. No. 68;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 LEIANSKDPNNMS 16  
Db 107 VDOIQAQIKOPDVS 121

RESULT 18  
US-08-430-024-2  
; Sequence 2, Application US/08430024  
; Patent No. 5644030  
; GENERAL INFORMATION:  
; APPLICANT: Faulmann, Ervin L  
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF  
; TITLE OF INVENTION: AN IGA BINDING PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,024  
; FILING DATE: 27-APR-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/677,209  
; FILING DATE: 29-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: BL-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-024-2

Query Match 38.6%; Score 39; DB 1; Length 396;  
Best Local Similarity 50.0%; Pred. No. 78;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;  
OY 4 EIAANSKDPNNMS 19  
Db 50 EVAA--KPPSMAQTD 63

RESULT 19  
US-08-782-009-2  
; Sequence 2, Application US/08782009  
; Patent No. 5714334  
; GENERAL INFORMATION:  
; APPLICANT: Faulmann, Ervin L  
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF  
; TITLE OF INVENTION: AN IGA BINDING PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,009  
FILING DATE: 07-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/430,024  
FILING DATE: 27-APR-1995  
APPLICATION NUMBER: US 07/677,209  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: BL-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 396 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-782-009-2

Query Match 38.6%; Score 39; DB 1; Length 396;  
Best Local Similarity 50.0%; Pred. No. 78;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

OY 4 EIAANSKDPNNMS 19  
Db 50 EVAA--KPPSMAQTD 63

RESULT 20  
US-09-017-302-2  
; Sequence 2, Application US/09017302  
; Patent No. 6075128  
; GENERAL INFORMATION:  
; APPLICANT: Faulmann, Ervin L  
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF  
; TITLE OF INVENTION: AN IGA BINDING PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/017,302  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/782,009  
FILING DATE: 07-JAN-1997  
APPLICATION NUMBER: US 08/430,024  
FILING DATE: 27-APR-1995  
APPLICATION NUMBER: US 07/677,209  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: BU-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 396 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-017-302-2

Query Match 38.6%; Score 39; DB 3; Length 396;  
Best Local Similarity 50.0%; Pred. No. 78;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

QY 4 EIAANSKDFPNMSETD 19  
|:| | :|:|:|  
DB 50 EVAA--KPYPSMAQTD 63

RESULT 21  
US-08-242-932-2  
Sequence 2, Application US/08242932  
Patent No. 5595740  
GENERAL INFORMATION:  
APPLICANT: Brady, L. Jeannine  
TITLE OF INVENTION: Cloning of No. 5595740-1GA Fc Binding Forms of  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,932  
FILING DATE: 16-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI42  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-372-5800  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 984 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-242-932-2

Query Match 38.6%; Score 39; DB 1; Length 984;  
Best Local Similarity 50.0%; Pred. No. 2,3e+02;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

QY 4 EIAANSKDFPNMSETD 19  
|:| | :|:|:|  
DB 13 EVAA--KPYPSMAQTD 26

RESULT 22  
US-08-714-481-2  
Sequence 2, Application US/08714481  
Patent No. 5766606  
GENERAL INFORMATION:  
APPLICANT: Brady, L. Jeannine  
TITLE OF INVENTION: Cloning of No. 5766606-1GA Fc Binding Forms of  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,481  
FILING DATE: 16-SEP-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/242,932  
FILING DATE: 16-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI42  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 984 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-714-481-2

Query Match 38.6%; Score 39; DB 1; Length 984;  
Best Local Similarity 50.0%; Pred. No. 2,3e+02;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

QY 4 EIAANSKDFPNMSETD 19  
|:| | :|:|:|  
DB 13 EVAA--KPYPSMAQTD 26

RESULT 23  
PCT-US95-06111-2  
Sequence 2, Application PC/TUS9506111  
GENERAL INFORMATION:

APPLICANT: Street address: 186 Grinter Hall  
APPLICANT: City: Gainesville  
APPLICANT: State/Province: Florida  
APPLICANT: Country: US  
APPLICANT: Postal code/zip: 32611  
APPLICANT: Phone number: 904-392-8929  
APPLICANT: Fax number: 904-392-6600  
TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of  
TITLE OF INVENTION: the Group B Streptococcal Beta Antigens  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06111  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/242,932  
FILING DATE: 16-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 984 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-06111-2

Query Match 38.6%; Score 39; DB 5; Length 984;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

OY 4 EIANSKDFPMSETD 19  
|:| | | :| | :| |  
DB 13 EVAA--KPYPSMAQTD 26

RESULT 24  
US-08-923-992A-8  
Sequence 8 Application US/08923992A  
Patent No. 6280738  
GENERAL INFORMATION:  
APPLICANT: Tai, Joseph Y.  
APPLICANT: Blake, Milan S.  
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B  
TITLE OF INVENTION: Streptococcal Beta Antigens  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,992A  
FILING DATE: 05-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/024,707  
FILING DATE: 06-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1098 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-923-992A-8

Query Match 38.6%; Score 39; DB 4; Length 1098;  
Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

OY 4 EIANSKDFPMSETD 19  
|:| | | :| | :| |  
DB 19 EVAA--KPYPSMAQTD 32

RESULT 25  
US-08-923-992A-4  
Sequence 4 Application US/08923992A  
Patent No. 6280738  
GENERAL INFORMATION:  
APPLICANT: Tai, Joseph Y.  
APPLICANT: Blake, Milan S.  
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B  
TITLE OF INVENTION: Streptococcal Beta Antigens  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,992A  
FILING DATE: 05-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/024,707  
FILING DATE: 06-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1104 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-923-992A-4

Query Match 38.6%; Score 39; DB 4; Length 1104;  
Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

OY 4 EIAANSKDPNMSFTD 19  
|:| | :|:| |  
Db 19 EVAA--KPYPSMAQTD 32

Search completed: June 16, 2002, 00:03:20  
Job time: 8534 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:22 ; Search time 108.75 Seconds  
(Without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-432a-45

Perfect score: 101  
Sequence: 1 DLEIANSKDFPNMSETDL 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 60 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	2227	1 GNNYHM	genome polyprotein
2	101	100.0	2227	1 GNNYHR	genome polyprotein
3	101	100.0	2227	1 GNNYMK	genome polyprotein
4	101	100.0	2227	1 GNNYHB	genome polyprotein
5	98	97.0	2230	1 GNNYSA	genome polyprotein
6	50	49.5	736	2 D90574	hypothetical prote
7	49	48.5	400	2 F88931	protein R1G11.1 [
8	45.5	45.0	930	2 A84668	Argonaute (AGO)-1
9	45.5	45.0	1621	2 T15264	hypothetical prote
10	44	43.6	740	1 FOLJHD	99g polyprotein -
11	44	43.6	927	2 T38127	phosphoprotein - f
12	44	43.6	1046	2 S67786	hypothetical prote
13	43.5	43.1	754	1 BABOH	peptide-aspartate
14	43.5	43.1	1451	2 S65571	pattern formation
15	43.5	43.1	5107	2 T29144	partial CDS - Caen
16	43	42.6	152	2 A64610	hypothetical prote
17	43	42.6	152	2 A71904	hypothetical prote
18	43	42.6	231	2 H70407	probable di-trans,
19	43	42.6	285	2 B87426	rhodanese family p
20	43	42.6	454	2 T02100	hypothetical prote
21	43	42.6	598	2 H71336	probable cell divi
22	43	42.6	1072	2 A84112	alkaline amylopull
23	43	42.6	1201	2 T08603	kinesin-related pr
24	42.5	42.1	913	2 T15278	hypothetical prote
25	42.5	42.1	4717	2 T41581	hypothetical prote
26	42	41.6	289	2 T23342	hypothetical prote
27	42	41.6	365	2 C81050	cytochrome c oxida
28	42	41.6	365	2 F81826	probable cytochrom
29	42	41.6	538	2 G95015	ABC transporter, A

30	42	41.6	538	2 A97889	hypothetical prote
31	42	41.6	1112	2 T47784	hypothetical prote
32	42	41.6	1946	2 A61449	hypothetical prote
33	41.5	41.1	236	2 A81283	probable periplasm
34	41.5	41.1	694	2 S41868	DNA-directed RNA p
35	41.5	41.1	1378	2 A81393	hypothetical prote
36	41	40.6	94	2 D70245	hypothetical prote
37	41	40.6	108	2 AF2042	hypothetical prote
38	41	40.6	133	2 A29174	ciostriatin (EC 3.
39	41	40.6	139	2 T26224	hypothetical prote
40	41	40.6	230	2 T45754	hypothetical prote
41	41	40.6	262	2 E89760	conserved hypotet
42	41	40.6	292	2 F66431	hypothetical prote
43	41	40.6	295	2 H83642	probable 2-OH-laur
44	41	40.6	346	2 G64182	lipopolysaccharide
45	41	40.6	368	2 G97281	hypothetical prote
46	41	40.6	376	2 A48681	placental chromin
47	41	40.6	387	2 F72416	alcohol dehydrogen
48	41	40.6	439	2 C32812	repc protein - Agr
49	41	40.6	493	2 T31794	hypothetical prote
50	41	40.6	526	2 S35190	ciostriatin (EC 3.
51	41	40.6	528	2 JN0619	glucuronosyltransf
52	41	40.6	548	2 B82486	hemolysin secretio
53	41	40.6	548	2 S15910	hemolysin-releasin
54	41	40.6	657	2 E81119	tail fibre protein
55	41	40.6	657	2 E81901	probable phage tai
56	41	40.6	686	2 C83734	beta-hexosaminid
57	41	40.6	776	2 T09938	hypothetical prote
58	41	40.6	776	2 T08913	cell division prot
59	41	40.6	1072	2 A66827	hypothetical prote
60	41	40.6	1111	2 T01239	hypothetical prote

## ALIGNMENTS

RESULT 1  
GNNYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:COHEN, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA5465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-401/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-960/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: core protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DLEIANSKDFPNMSETDL 20

```
Db      931 DLEEIANSKDPFNMSETDL 950
|||||
RESULT  2
GNMYH
genome polypotein - human hepatitis A virus
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro
NA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03903
R:NaJarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985
A:Title: Primary structure and gene organization of human hepatitis A virus.
A:Reference number: A03903; MUID:85130549
A:Accession: A03903
A:Molecule type: genomic RNA
A:Residues: 1-2227 <NM>
C:Cross-references: GB:K02990; NID:9329596; PIDN:AAA45472.1; PID:9329597
C:Superfamily: hepatitis A virus genome polypotein
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans
F:1-245/Product: coat protein 1A #status predicted <C1A>
F:1245-491/Product: coat protein 1B #status predicted <C1B>
F:492-836/Product: coat protein 1C #status predicted <C1C>
F:837-980/Product: core protein 2A #status predicted <C2A>
F:981-1076/Product: core protein 2B #status predicted <C2B>
F:1077-1422/Product: core protein 2C #status predicted <C2C>
F:1423-1484/Product: protein 3A #status predicted <C3A>
F:1485-1507/Product: protein 3B #status predicted <C3B>
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match      100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      931 DLEEIANSKDPFNMSETDL 950
|||||
RESULT  3
GNMYH
genome polypotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro
NA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A94149; A25914; A94508
R:Cohen, J.I.; Rosenblum, B.; Tlichurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with
A:Reference number: A94149; MUID:87175701
A:Accession: A94149
A:Molecule type: genomic RNA
A:Residues: 1-2227 <CO>
A:Cross-references: EMBL:M16632; NID:9329594; PIDN:AAA45471.1; PID:9329595
A:Note: submitted to GenBank, August 1987
C:Superfamily: hepatitis A virus genome polypotein
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans
F:1-245/Product: coat protein 1A #status predicted <C1A>
F:1245-491/Product: coat protein 1B #status predicted <C1B>
F:492-836/Product: coat protein 1C #status predicted <C1C>
F:837-980/Product: core protein 2A #status predicted <C2A>
F:981-1076/Product: core protein 2B #status predicted <C2B>
F:1077-1422/Product: core protein 2C #status predicted <C2C>
F:1423-1484/Product: protein 3A #status predicted <C3A>
F:1485-1507/Product: protein 3B #status predicted <C3B>
F:1508-1678/Product: protein 3C #status predicted <C3C>
F:1679-2227/Product: protein 3D #status predicted <C3D>
```

```
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match      100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      931 DLEEIANSKDPFNMSETDL 950
|||||
RESULT  4
GNMYH
genome polypotein - human hepatitis A virus (strain MBB)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
Vpg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D
C:Species: human hepatitis A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-May-1996
C:Accession: J50303
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Klein, R.; Wimmer, E.; Deinhard
Virus Res. 6, 153-171, 1987
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso
A:Reference number: J50303; MUID:88045071
A:Accession: J50303
A:Molecule type: genomic RNA
A:Residues: 1-2227 <PAU>
C:Cross-references: EMBL:M20273
C:Superfamily: hepatitis A virus genome polypotein
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h
F:1-23/Product: coat protein 1A #status predicted <VP4>
F:24-246/Product: coat protein 1B #status predicted <VP2>
F:247-491/Product: coat protein 1C #status predicted <VP3>
F:492-836/Product: coat protein 1D #status predicted <VP1>
F:837-980/Product: core protein 2A #status predicted <P2A>
F:981-1108/Product: core protein 2B #status predicted <P2B>
F:1109-1438/Product: core protein 2C #status predicted <P2C>
F:1439-1496/Product: protein 3A #status predicted <P3A>
F:1497-1519/Product: genome-linked protein Vpg #status predicted <VPG>
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match      100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      931 DLEEIANSKDPFNMSETDL 950
|||||
RESULT  5
GNMYSA
genome polypotein - simian hepatitis A virus (strain AGM-27)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
C:Species: simian hepatitis A virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jun-2000
C:Accession: A30470; S04885; S03965
R:Tzarev, S.A.
submitted to JIPID, April 1991
A:Reference number: A30470
A:Accession: A30470
A:Molecule type: genomic RNA
A:Residues: 1-2230 <TSA>
A:Cross-references: GB:D00924; NID:9222597; PIDN:BAA00766.1; PID:9222598
F:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Tlichurst, J.; Purcell, R.H.
J. Gen. Virol. 72, 1677-1683, 1991
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure
A:Reference number: J01080; MUID:91311420
A:Contents: annotation
```

A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjapardize, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhnik  
submitted to the EMBL Data Library, May 1989  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:961971; PIDN:CAA33490.1; PID:9930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjapardize, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhnik  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 97.0%; Score 98; DB 1; Length 2230;  
Best Local Similarity 95.0%; Pred. No. 6.6e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETANSKDFPNMSFTDL 20  
|||:|||||:|||||:  
DB 935 DLEETANSKDFPNMSFTDL 954

RESULT 6  
D90574  
hypothetical protein MYPU\_5000 [imported] - Mycoplasma pulmonis (strain UAB CTRP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: D90574  
R:Chambud, I.; Heiligg, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: D90574  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-736 <KUR>  
A:Cross-references: GB:AL445566; PID:914089914; PIDN:CAC13673.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTRP  
C:Genetics:  
A:Gene: MYPU\_5000  
A:Genetic code: SGC3

Query Match 49.5%; Score 50; DB 2; Length 736;  
Best Local Similarity 55.0%; Pred. No. 8.3;  
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 DLEETANSKDFPNMSFTDL 20  
|||:|||||:|||||:  
DB 218 DLYSIYNEKDAPEISEDL 237

RESULT 7  
F88931  
protein R1G11.1 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: F88931  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1996  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-  
A:Accession: F88931  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-400 <STO>  
A:Cross-references: GB:chr\_V; PIDN:AAC69076.1; PID:92384851; GSPDB:GN00023; CESP:R1IG  
C:Genetics:  
A:Gene: R1G11.1  
A:Map position: 5

Query Match 48.5%; Score 49; DB 2; Length 400;  
Best Local Similarity 47.4%; Pred. No. 5.9;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LLEETANSKDFPNMSFTDL 20  
|||:|||||:|||||:  
DB 164 LENTSNKDFPHSKEDV 182

RESULT 8  
A84668  
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A84668  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: A84668  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-930 <STO>  
A:Cross-references: GB:AE02093; NID:93885334; PIDN:AAC77862.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g27040  
A:Map position: 2  
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 45.0%; Score 45.5; DB 2; Length 930;  
Best Local Similarity 42.3%; Pred. No. 57;  
Matches 11; Conservative 4; Mismatches 2; Indels 9; Gaps 1;

QY 2 LLEETANSKDF-----PNMSFT 18  
|||:|||||:|||||:  
DB 153 LEEVATSKDFVSRANGSPNGES 178

RESULT 9  
T15264  
hypothetical protein F59E12.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15264  
R:Johnson, D.  
submitted to the EMBL Data Library, May 1997  
A:Description: The sequence of C. elegans cosmid F59E12.  
A:Reference number: Z18318  
A:Accession: T15264  
A:Status: preliminary; translated from GB/EMBL/JDBJ





A:Residues: 1-754 <JIA>  
A:Cross-references: EMBL:M91213; NID:g162693; PIDN:AAA03563.1; PID:g162694  
A:Experimental source: Brain  
A:Note: sequence extracted from NCBI backbone (NCBIP:108534)  
R:Man, O.; Vandusen, W.J.; Petroki, C.J.; Garaky, V.M.; Stern, A.M.; Friedman, P.A.  
J. Biol. Chem. 266, 14004-14010, 1991  
A:Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.  
A:Reference number: A39470; MUID:91310689  
A:Accession: A39470  
A:Molecule type: Protein  
A:Residues: 289-328 <MAN>  
A:Accession: B39470  
A:Molecule type: Protein  
A:Residues: 615, 'X', 617-630, 'XX', 633-634, 'X', 636, 'XX', 639-641 <MA>  
A:Accession: C39470  
A:Molecule type: Protein  
A:Residues: 311-347, 'X', 349, 'X', 351-373, 'X', 375-379, 'X', 381-382 <MA>  
C:Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the  
C:Comment: Aspartic acid and asparagine residues in the EGF homology domain of certain F  
C:Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology  
C:Keywords: glycoprotein; intracellular; oxidoreductase; transmembrane protein  
F:2-56/Domain: Intracellular #status predicted <INC>  
F:57-78/Domain: transmembrane #status predicted <TRM>  
F:289-754/Product: peptide-aspartate beta-dioxygenase, 56k form #status predicted <56K>  
F:311-754/Product: peptide-aspartate beta-dioxygenase, 52k form #status predicted <52K>  
F:337-370/Domain: tetratricopeptide repeat homology <TT1>  
F:371-404/Domain: tetratricopeptide repeat homology <TT2>  
F:13,96,466,702/Binding site: carbohydrate (Asn).(covalent) #status predicted

Query Match 43.18; Score 43.5; DB 1; Length 754;  
Best Local Similarity 55.68; Pred. No. 93;  
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;  
OY 1 DLEETANSKDFP-NMSE 17  
| 11: 1 111 111  
Db 82 DYEVLAIAKDFRNLSE 99

RESULT 14  
S65571  
Pattern formation protein GNOM - Arabidopsis thaliana  
N:Alternate names: EMB30 protein  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 07-May-1999  
C:Accession: S65571; S65572  
R:Busch, M.; Mayer, U.; Juergens, G.  
Mol. Gen. Genet. 250, 681-691, 1996  
A:Title: Molecular analysis of the Arabidopsis pattern formation gene GNOM: gene structure  
A:Reference number: S65571; MUID:96204508  
A:Accession: S65571  
A:Molecule type: DNA  
A:Residues: 1-1451 <BUS>  
A:Cross-references: EMBL:U03433; NID:g1209632; PID:g1209633  
A:Accession: S65572  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-110, 'I', 112-866, 'G', 868-1451 <BUS>  
A:Cross-references: EMBL:U36432; NID:g1209630; PID:g1209631  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1995  
C:Genetics:  
A:Gene: GNOM; EMB30  
A:Introns: 246/3

Query Match 43.18; Score 43.5; DB 2; Length 1451;  
Best Local Similarity 63.28; Pred. No. 2e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
OY 1 DLEETANS-KDFPMSET 18  
| 1111 111 11  
Db 1302 DLEETANSKDFRNMECT 1320

RESULT 15  
T29144  
partial CDS - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T29144  
R:Pauley, A.; Gatlung, S.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid K11C4.  
A:Reference number: 220577  
A:Accession: T29144  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5107 <PAU>  
A:Cross-references: EMBL:U64854; PIDN:AAB18318.1; GSPDB:GN00023; CESP:unc-68  
A:Experimental source: strain Bristol N2; clone K11C4  
C:Genetics:  
A:Gene: CESP:unc-68  
A:Map position: 5  
A:Introns: 27/1; 64/3; 92/3; 127/1; 158/2; 1222/2; 1300/2; 1347/2; 1391/1; 1419/3; 15  
/3; 3269/2; 3313/2; 3466/1; 3519/3; 3615/3; 3629/3; 3658/2; 3710/1; 3741/3; 3779/2; 3  
C:Superfamily: ryanodine receptor; transcription initiation factor sigma region 1 hom

Query Match 43.18; Score 43.5; DB 2; Length 5107;  
Best Local Similarity 43.58; Pred. No. 8.4e+02;  
Matches 10; Conservative 4; Mismatches 6; Indels 3; Gaps 1;  
OY 1 DLEETANS---KDFPMSETDL 20  
| 11 11 111 111  
Db 3377 DLEETANNNTMYSDPNVYDVL 3399

RESULT 16  
A64610  
Hypothetical protein HP0721 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 20-Jun-2000  
C:Accession: A64610  
R:Tom, J.F.; White, O.; Kellavag, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKe  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey,  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467  
A:Accession: A64610  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-152 <TCM>  
A:Cross-references: GB:AE000585; GB:AE000511; NID:g2313845; PIDN:AAD07777.1; PID:g231  
C:Superfamily: Helicobacter pylori hypothetical protein jhp0118

Query Match 42.68; Score 43; DB 2; Length 152;  
Best Local Similarity 50.08; Pred. No. 18;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
OY 5 TAANSKDFPMSETDL 20  
| 1111 11 11  
Db 16 VALNAKDFSKTSDEDL 31

RESULT 17  
A71904  
Hypothetical protein jhp0658 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Jun-2000  
C:Accession: A71904  
R:Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, O.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*  
 A:Reference number: A71800; MUID:99120557  
 A:Accession: A71904  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-152 <ARN>  
 A:Cross-references: GB:AE001498; GB:AE001439; NID:g415521; PIDN:AAD06236.1; PID:g415521  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp0658  
 C:Superfamily: Helicobacter pylori hypothetical protein jhp0118

Query Match	42.6%	Score 43	DB 2	Length 152
Best Local Similarity	50.0%	Pred. No. 18		
Matches	8	Conservative	3	Mismatches 5; Indels 0; Gaps 0;
Oy	5	IAANSKDFPMSEDTL	20	
	:		1: 11	
Db	16	VALNAKDFSKTSEDTL	31	

RESULT 18  
 H70407  
 probable di-trans,poly-cis-decaprenyl:sttransferase (EC 2.5.1.31) ag\_1248 - Aquifex aeolicus  
 N:Alternate names: decaprenyl diphosphate synthase  
 C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 24-Sep-1999  
 C:Accession: H70407  
 V:Ridekett, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196566  
 A:Accession: H70407  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-231 <ADP>  
 A:Cross-references: GB:AE000730; MTD:g2983674; PIDN:AAC07254.1; PID:g2983689; GB:AE00065  
 A:Experimental source: strain VF5  
 C:Genetics:  
 A:Gene: ag\_1248  
 C:Superfamily: conserved hypothetical protein YBR002c  
 C:Keywords: transferase

Query Match	42.6%	Score 43	DB 2	Length 231
Best Local Similarity	50.0%	Pred. NO. 29		
Matches	7	Conservative	5	Mismatches
			2	Indels
			0	Gaps
			0	
QY	2	LEETIANSKDFPNM	15	
	:		:::	
DB	112	MEETESDSKDFPNL	125	

RESULT 19  
B87426  
rhodanese family protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87426  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Emdolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11359647  
A:Accession: B87426  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <STO>  
A:Cross-references: GB:AE005673; NID:G13422788; PIDN:AAK23406.1; GSPDB:GN00148

C/Genetics: A/Gene: CC1425 C/Superfamily: thiosulfate sulfurtransferase

Query March	42.6%	Score 43	DB 2	Length 285
Best Local Similarity	53.3%	Pred. NO. 36		
Matches 8	Conservative	3	Mismatches 4	Indels 0
Gaps 0				
QY	1	DLEETANSKDEPPNM	15	
	::		:	
DB	55	DIDEIADSDTLPHM	69	

RESULT 20  
T02100  
hypothetical protein Atg41190 [imported] - Arabidopsis thaliana  
N/Alternate names: hypothetical protein T3K9.4  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 26-Feb-1999 #sequence,revision 26-Feb-1999. #text\_change 16-Feb-2001  
C/Accession: T02100; H84838  
R/Rounsailey, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke  
Submitted to the EMBL Data Library, February 1999  
A/Description: Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence.  
A/Reference number: Z14570  
A/Accession: T02100  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-454 <ROU>  
A/Cross-references: EMBL:AC004261; NID:g3402695; PID:g3402699  
A/Experimental source: cultivar Columbia  
R/Lin, X.; Kaul, S.; Rounsailey, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C. Y  
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,  
Eugene, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487  
A/Accession: H84838  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-454 <STO>  
A/Cross-references: GB:AE002093; NID:g3402699; PID:MAD11993.1; GSPDB:GN00139  
C/Genetic:  
A:Gene: T3K9.4; Atg41190  
A/Map position: 2  
A/Introns: 72/2; 156/1; 227/3; 235/3; 296/1; 337/1; 392/2; 407/1; 428/1

Query Match	42.6%	Score 43	DB 2	Length 454
Best Local Similarity	47.4%	Pred. No. 62		
Matches	9	Conservative	3	Mismatches 7
				Indels 0
				Gaps 0
QY	1	DLEETIANSKDEPNNSSTD	19	
		:::		
		:::		
		:::		
DB	21	DNEDEEDNSSKYEENDSTD	39	

RESULT 21  
H71336  
probable cell division protein - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
C:Accession: H71336  
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, F.  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uetzbach, T.  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: H71336  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-598 <COL>





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:10:11 ; Search time 59.43 Seconds

(without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432a-45

Perfect score: 101

Sequence: 1 DLEETIANSKDEPNMSETDL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## ALIGNMENTS

34	39	38.6	396	1	L1DD_ECOLI	P33232	eschericchia
35	39	38.6	396	1	YJEB_YEAST	P47049	saccharomyc
36	39	38.6	438	1	GUX3_AGAB1	P49075	agaricus b1
37	39	38.6	474	1	CC2H_CRIFA	001917	crithidia f
38	39	38.6	642	1	Y622_YEAST	P53235	saccharomyc
39	39	38.6	660	1	DNAK_CHLPH	P27542	chlamydia p
40	39	38.6	673	1	VID3_AGR5	P18593	saccharomyc
41	39	38.6	674	1	YB66_YEAST	P27951	streptococc
42	39	38.6	1164	1	BAG_SYRAG	P53243	saccharomyc
43	39	38.1	794	1	Y62A_YEAST	P17422	bacteroides
44	38.5	38.1	860	1	CLPB_BACNO	P55076	barbus inte
45	38	37.6	116	1	B2MG_BARIN	P07151	tattus notv
46	38	37.6	119	1	B2MG_RAT	Q28747	cercosporu
47	38	37.6	205	1	IL6_ORCRO	P46550	cercosporu
48	38	37.6	212	1	IL6_CERTO	P05321	homo sapien
49	38	37.6	212	1	IL6_HUMAN	P79341	macaca fasc
50	38	37.6	212	1	IL6_MACFA	P51494	macaca mula
51	38	37.6	212	1	IL6_MACMU	P76369	homo sapien
52	38	37.6	228	1	CUDA_HUMAN	O94db6	homo sapien
53	38	37.6	421	1	Y416_RICPR	P83110	homo sapien
54	38	37.6	452	1	HRA3_HUMAN	P37093	vibrio chol
55	38	37.6	503	1	GSPE_VIBCH	P43975	haemophilus
56	38	37.6	551	1	Y275_HAETN	P45962	caenorhabd1
57	38	37.6	598	1	K1P3_CAEL	P35527	homo sapien
58	38	37.6	622	1	YBBD_HUMAN	P40406	baecillus su
59	38	37.6	642	1	YBBD_BACSV	P53835	saccharomyc
60	38	37.6	661	1	YNI9_YEAST		

Result No.	Score	Query Match	Length	DB ID	Description		
1	101	100.0	2226	1	POLG_HPVA4	P26581	hepatitis a
2	101	100.0	2226	1	POLG_HPVA8	P26582	hepatitis a
3	101	100.0	2227	1	POLG_HPVAH	P08617	hepatitis a
4	101	100.0	2227	1	POLG_HPVAL	P08441	hepatitis a
5	101	100.0	2227	1	POLG_HPVAH	P13901	hepatitis a
6	98	97.0	2230	1	POLG_HPVAH	P14553	slman hepa
7	97	96.0	2226	1	POLG_HPVA2	P26580	hepatitis a
8	46	45.5	194	1	Y041_THEAC	P57674	thermoplasm
9	44	43.6	740	1	GAG_SHRVA	P21411	squirtel mo
10	44	43.6	927	1	CC15_SCHPO	Q09822	schizosacch
11	43.5	43.1	754	1	ASPH_BOVIN	Q20856	bos taurus
12	43.5	43.1	1451	1	EM30_AARTH	Q4510	archidopsis
13	43	42.6	231	1	UPPS_AOUAE	O67291	aquifex aeo
14	43	42.6	350	1	40MT_CORJA	O91e15	coplus japo
15	41.3	41.1	1378	1	RPOB_CAME	Q46124	campylobact
16	41	40.6	346	1	RPAF_HAETN	P45042	haemophilus
17	41	40.6	376	1	PT16_HUMAN	P35237	homo sapien
18	41	40.6	526	1	CLOS_CLOHI	P09870	clostridium
19	41	40.6	528	1	UNB4_HUMAN	P06133	homo sapien
20	41	40.6	548	1	H1XB_VIBCH	P13492	vibrio chol
21	40	39.6	283	1	RUAP_SOYBN	P33657	glycine max
22	40	39.6	381	1	L1DD_HAETN	P46544	haemophilus
23	40	39.6	547	1	NLTP_HUMAN	P22307	homo sapien
24	40	39.6	824	1	NSFH_CAEL	O94392	caenorhabd1
25	40	39.6	1087	1	ELI3_HUMAN	O94212	homo sapien
26	39.5	39.1	244	1	P29_MYCPN	P73370	mycoplasma
27	39.5	39.1	521	1	VI2_HPV04	Q07084	human papil
28	39.5	39.1	552	1	YB81_MYCTU	Q05684	mycobacteri
29	39	38.6	221	1	YSC1_YERPE	Q00928	yersinia pe
30	39	38.6	223	1	YSC1_YEREN	Q01253	yersinia en
31	39	38.6	253	1	PCRB_MERTJ	O58647	methanococc
32	39	38.6	320	1	Y149_MYCNA	P75583	mycoplasma
33	39	38.6	394	1	UXUA_SALTY	P43668	salmonella

RESULT 1  
POLG\_HPVA4  
ID POLG\_HPVA4 STANDARD: PRT: 2226 AA.

AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide (Contains: Coat proteins VP1 to VP4; Core proteins P2A to P2C; Probable proteins P3A to P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)).  
OS Hepatitis A virus (strain 43c).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxId=12095;  
RN [1]  
RP MEDLINE=91162758; PubMed=1705995;  
RX Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination.";  
RT J. Virol. 65:2056-2065(1991).  
RL -i- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -i- P1M: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -i- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL collaboration - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL: M59809: AAA45469.1: -  
DR MEROPS: C03.005: -  
DR InterPro: IPR000605: RNA\_helicase.  
DR InterPro: IPR001205: RNA\_pol\_P3D.

DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KM Polyprotein; Coat protein; Core protein; Transferase;  
FT CHAIN RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 CORE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CAB0B09BF75 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLEEIAANSKDPFNMSETDL 20  
Db 931 DLEEIAANSKDPFNMSETDL 950  
|||||

RESULT 2  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
ID POLG\_HPAV8  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48).  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxId=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination".  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC  
CC EMBL: M59808; AAA5467.1; -  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3B.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KM Polyprotein; Coat protein; Core protein; Transferase;  
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).

FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 CORE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8DB8 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLEEIAANSKDPFNMSETDL 20  
Db 931 DLEEIAANSKDPFNMSETDL 950  
|||||

RESULT 3  
POLG\_HPAV8 STANDARD; PRT; 2227 AA.  
ID POLG\_HPAV8  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48).  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxId=12098;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WILD TYPE;  
RX MEDLINE=87061253; PubMed=3023706;  
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
RA Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus:  
RT comparison with different strains of hepatitis A virus and other  
RT picornaviruses.";  
RL J. Virol. 61:50-59(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATTENUATED;  
RX MEDLINE=87175701; PubMed=3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,  
RA Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RT comparison with wild-type virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RN [3]  
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE=85166289; PubMed=2984684;  
RA Baroudy B.M., Ticehurst J.R., Mele T.A., Matzel J.V. Jr.,  
RA Purcell R.H., Feinstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RT proteins and RNA polymerase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: STRAIN HM-175/77 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M14114; AAA45475.1; -  
DR EMBL; M14707; AAA45465.1; -  
DR EMBL; M14707; AAA45466.1; ALT\_INT.  
DR EMBL; M16632; AAA45471.1; -  
DR PIR; A25981; GNNYHM.  
DR PIR; A25914; GNNYMK.  
DR PIR; A03905; A03905.  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT VARIANT 77 77  
FT VARIANT 764 821  
FT VARIANT 1052 1052  
FT VARIANT 1062 1062  
FT VARIANT 1118 1118  
FT VARIANT 1151 1151  
FT VARIANT 1163 1163  
FT VARIANT 1277 1277  
FT VARIANT 1500 1500  
FT VARIANT 1805 1805  
FT VARIANT 1930 1930  
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLEETANSKDFPNMSETDL 20  
Db 931 DLEETANSKDFPNMSETDL 950

RESULT 4  
POLG\_HPAVL STANDARD; PRT; 2227 AA.  
ID POLG\_HPAVL  
AC P06441;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12099;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=85190549; PubMed=2986127;  
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; K02990; AAA45472.1; -  
DR PIR; A03903; GNNYHR.  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1076  
FT CHAIN 1077 1422  
FT CHAIN 1423 1484  
FT CHAIN 1485 1507  
FT CHAIN 1508 1678  
FT CHAIN 1679 2227  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DLEETANSKDFPNMSETDL 20  
Db 931 DLEETANSKDFPNMSETDL 950

RESULT 5  
POLG\_HPAVM STANDARD; PRT; 2227 AA.  
ID POLG\_HPAVM  
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MB9).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wessel T., Klein R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
RT virus (isolate MB9).";

```

RA Virus Res. 8:153-171(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M20273; AAA5474.1; -.
CC PIR: J50303; GNNYHB.
CC DR MEROPS: C03.005; -.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC DR Pfam: PF00680; RNA_dep_RNA_pol: 1.
CC KM Polyprotein: Coat protein: Core protein: Transferase:
CC KW RNA-directed RNA polymerase: Hydrolase: Thiol protease.
CC FT CHAIN 1 33
CC FT CHAIN 24 245
CC FT CHAIN 246 491
CC FT CHAIN 492 836
CC FT CHAIN 837 980
CC FT CHAIN 981 1087
CC FT CHAIN 1088 1432
CC FT CHAIN 1433 1496
CC FT CHAIN 1497 1519
CC FT CHAIN 1520 1738
CC FT CHAIN 1739 2227
CC SO SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 8.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEIANSKDPFNMSETDL 20
DB 931 DLEIANSKDPFNMSETDL 950

RESULT 6
POLG_HPAVS STANDARD; PRT; 2230 AA.
AC P14553;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
DE Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsarev S.A., Emerson S.O., Balyan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RA "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RA structure and growth in cell culture with other HAV strains.";
RA J. Gen. Virol. 72:1677-1683(1991).
RN [2]
RX SEQUENCE OF 1750-2164 FROM N.A.
RP MEDLINE=89232168; PubMed=2541023;
RA Balyan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
```

```

RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RT "Variations in genome fragments coding for RNA polymerase in human
RT and simian hepatitis A viruses.";
RL FEBS Lett. 247:425-428(1989).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D00924; BAA00766.1; -.
CC DR EMBL: X15461; CAA33490.1; -.
CC DR PIR: A30470; GNNYSA.
CC DR PIR: S04885; S04885.
CC DR MEROPS: C03.005; -.
CC DR InterPro: IPR000605; RNA_helicase.
CC DR InterPro: IPR001205; RNA_pol_P3D.
CC DR Pfam: PF00680; RNA_dep_RNA_pol: 1.
CC DR Pfam: PF00910; RNA_helicase: 1.
CC KM Polyprotein: Coat protein: Core protein: Transferase:
CC KW RNA-directed RNA polymerase: Hydrolase: Thiol protease.
CC FT CHAIN 1 27
CC FT CHAIN 28 249
CC FT CHAIN 250 495
CC FT CHAIN 496 795
CC FT CHAIN 796 984
CC FT CHAIN 985 1091
CC FT CHAIN 1092 1426
CC FT CHAIN 1427 1498
CC FT CHAIN 1499 1521
CC FT CHAIN 1522 1741
CC FT CHAIN 1742 2230
CC SO SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match 97.0%; Score 98; DB 1; Length 2230;
Best Local Similarity 95.0%; Pred. No. 2.5e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEIANSKDPFNMSETDL 20
DB 935 DLEIANSKDPFNMSETDL 954

RESULT 7
POLG_HPAV2 STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
DE Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12094;
RN [1]
RP SEQUENCE FROM N.A.
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Felstone S.M.,
RA Cromean T., Jensen R.W.;
RA "Antigenic and genetic variation in cytopathic hepatitis A virus
RA variants arising during persistent infection: evidence for genetic
```



RT recombination.".  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M59810: AAA45468.1: -.  
DR MEROPS: C03.005: -.  
DR InterPro: IPR000605: RNA\_helicase.  
DR InterPro: IPR001205: RNA\_pol\_P3D.  
DR Pfam: PF00680: RNA\_dep\_RNA\_pol: 1.  
DR Pfam: PF00910: RNA\_helicase: 1.  
KM Polypeptide: Coat protein; Core protein; Transferase;  
KM RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 CORE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;  
  
Query Match 96.0%; Score 97; DB 1; Length 2226;  
Best Local Similarity 95.0%; Pred. No. 3.6e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLEBIANSKDFPMNSETDL 20  
ID Y041\_THFAC STANDARD; PRT; 194 AA.  
AC P57674;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein TA0041.  
GN TA0041.  
OS Thermoplasma acidophilum.  
OC Archaea: Euryarchaeota: Thermoplasmatales: Thermoplasmaceae;  
OC Thermoplasmatales;  
OC NCBI\_TaxID=2303;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=DSM 1728;  
RX MEDLINE=20479972; Pubmed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermophilic scavenger Thermoplasma  
RT acidophilum.";  
RL Nature 407:508-513(2000).  
CC -1- SIMILARITY: BELONGS TO THE UPE0129 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: AL45063; GAC11190.1: -.  
DR InterPro: IPR002851; DUF133.  
DR Pfam: PF02000; DUF133: 1.  
DR Prodom: PD013236; DUF133: 1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 194 AA; 21854 MW; 6F7CC0D8449FF4F CRC64;  
  
Query Match 45.5%; Score 46; DB 1; Length 194;  
Best Local Similarity 50.0%; Pred. No. 3.1;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 LEEIANSKDFPMNSETD 19  
ID Y041\_THFAC STANDARD; PRT; 740 AA.  
AC P21411;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Gag polyprotein [Contains: Core protein p19; Core protein p16;  
DE Probable core protein p35; Probable core protein p10].  
GN Gag.  
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).  
OC Viruses; Retroviridae; Retroviridae; Betaretrovirus.  
OC NCBI\_TaxID=11856;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=89073750; Pubmed=3201749;  
RA Oda T., Ikeda S., Watanabe S., Hattushika M., Akiyama K.,  
RA Mitsunobu F.;  
RT "Molecular cloning, complete nucleotide sequence, and gene structure  
RT of the provirus genome of a retrovirus produced in a human  
RT lymphoblastoid cell line.";  
RL Virology 167:468-476(1988).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M23385; AAA66451.1: -.  
DR PIR: A31827; FOLJHD.  
DR InterPro: IPR003322; Gag\_p10.  
DR InterPro: IPR000721; Gag\_P24.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam: PF02337; Gag\_P10: 1.  
DR Pfam: PF00607; gag\_P24: 1.  
DR Pfam: PF00098; zf-CCHC: 1.  
DR SMART: SM00343; Znf\_C2HC: 1.  
KM Core protein; Polyprotein.  
FT CHAIN 1 163 CORE PROTEIN P19.  
FT CHAIN 164 318 CORE PROTEIN P16.  
FT CHAIN 319 648 PROBABLE CORE PROTEIN P35.  
FT CHAIN 649 740 PROBABLE CORE PROTEIN P10.  
SQ SEQUENCE 740 AA; 80543 MW; CC2503C6261221F5 CRC64;  
  
Query Match 43.6%; Score 44; DB 1; Length 740;  
Best Local Similarity 50.0%; Pred. No. 27;

Matches 10; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

OY 1 DLEIAA--NSKDFPMMSF 18  
 DB 214 DLEIAAQQNNPMDPOLNT 233

RESULT 10  
 CC15\_SCHPO STANDARD: PRT: 927 AA.  
 AC 009822; 014365;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN CDC15 OR SPAC20G8.05C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=95360987; PubMed=7634333;  
 RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,  
 RA Simanis V.;  
 RT "The S. pombe cdc15 gene is a key element in the reorganization of F-  
 actin at mitosis.";  
 RL Cell 82:435-444(1995).  
 RN [2]  
 RP REVISIONS TO N-TERMINUS.  
 RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,  
 RA Simanis V.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, FORMS A RING-LIKE STRUCTURE  
 WHICH CO-LOCALIZES WITH THE MEDIAL ACTIN RING. APPEARS TO MEDiate  
 CYTOSKELETAL REARRANGEMENTS REQUIRED FOR CYTOKINESIS. ESSENTIAL  
 FOR VIABILITY.  
 CC -1- DEVELOPMENTAL STAGE: PEAKS IN EARLY MITOSIS BEFORE SEPARATION.  
 CC -1- DOMAIN: THE N-TERMINAL REGION IS IN A COILED COIL STRUCTURE.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: SOME, TO S.POMBI: SPB11C11.02 AND SPAC7D4.02C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X86179; CAB0115.1; -  
 CC EMBL: 295334; CAB08599.2; -  
 DR HSSP: P07751; ITUD.  
 DR InterPro: IPR001060; FCH.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00611; FCH; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR SMART: SM00055; FCH; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KM Mitosis: Cytoskeleton: SH3 domain; Phosphorylation; Coiled coil.  
 FT DOMAIN 24 110 FCH.  
 FT DOMAIN 108 207 COILED COIL (POTENTIAL).  
 FT DOMAIN 866 927 SH3.

SQ SEQUENCE 927 AA; 102119 MW; PDCE7E0AAA3D247D CRC64;

Query Match 43.6%; Score 44; DB 1; Length 927;  
 Best Local Similarity 47.1%; Pred. No. 35;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 LDEIAANSKDFPMMSF 18  
 DB 69 LQELAASSADIREVGST 85

RESULT 11  
 ASPH\_BOVIN STANDARD: PRT: 754 AA.  
 AC 028056;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Asparyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-  
 DE hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-  
 DE dioxygenase).  
 GN ASPH.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 NX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver, and Brain;  
 RX MEDLINE=92332546; PubMed=1378441;  
 RA Jia S., Vandusen W.J., Diehl R.E., Kohl N.E., Dixon R.A.F.,  
 RA Elliston K.O., Stern A.M., Friedman P.A.;  
 RT "cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-  
 RT hydroxylase.";  
 RL J. Biol. Chem. 267:14322-14327(1992).  
 RN [2]  
 RP SEQUENCE OF 289-385 AND 615-641.  
 RC TISSUE=Liver;  
 RX MEDLINE=91310689; PubMed=1856229;  
 RA Wang O., Vandusen W.J., Petroski C.J., Garsky V.M., Stern A.M.,  
 RA Friedman P.A.;  
 RT "Bovine liver aspartyl beta-hydroxylase. Purification and  
 RT characterization.";  
 RL J. Biol. Chem. 266:14004-14010(1991).  
 CC -1- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN  
 CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF  
 CC PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) -  
 CC peptide 3-hydroxy-L-aspartate + succinate + CO(2).  
 CC -1- COFACTOR: IRON.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic  
 CC reticulum.  
 CC -1- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 289-754) OR 52 kDa (AA  
 CC 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M91213; AAA03563.1; -  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR; 2.  
 KM Oxidoreductase: Dioxxygenase; Iron; Transmembrane; Signal-anchor;  
 KM Endoplasmic reticulum.  
 FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 58 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)



CC isopentenyl diphosphate - diphosphate + di-trans-poly-cis-  
 CC undecaprenyl diphosphate.  
 CC -1- SIMILARITY: BELONGS TO THE UPP SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF000730; AAC07254.1; -  
 CC InterPro: IPR001441; UPP\_synth.  
 CC Pfam: PF01255; UPP\_synthetase; ..  
 CC ProDom: PD003461; UPP\_synth; 1.  
 CC PROSITE: PS01066; UPP\_SYNTHETASE; 1.  
 CC K M Transferrase; Cell division; Cell wall; Peptidoglycan synthesis;  
 CC Complete proteome.  
 CC SEQUENCE 231 AA; 27148 MW; 52B9DE0A442B80AE CRC64;  
 CC -----  
 CC Query Match 42.6%; Score 43; DB 1; Length 231;  
 CC Best Local Similarity 50.0%; Pred. No. 11;  
 CC Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 CC -----  
 CC Y 2 LEEIANSKDPNM 15  
 CC :||:|||||:  
 CC Db 112 MEELSDSKDFKNL 125  
 CC -----  
 CC RESULT 14  
 CC ID 4OMT\_COPUA STANDARD; PRT; 350 AA.  
 CC AC 09LE15;  
 CC DT 01-MAR-2002 (Rel. 41, Created)  
 CC DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 CC DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 CC DE 3'-hydroxy-N-methyl-(S)-coclaurine 4'-O-methyltransferase  
 CC (EC 2.1.1.16) (S-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine  
 CC 4'-O-methyltransferase) (4'-OMT).  
 CC DE Coplis japonica (Japanese goldfinch).  
 CC OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; endicotsyledons; Ranunculales;  
 CC OC Ranunculaceae; Coptis.  
 CC OX NCBI\_Taxid=3442;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 CC RX MEDLINE=20390108; PubMed=10811648;  
 CC RA Morishige T., Tsujita T., Yamada Y., Sato F.;  
 CC RT "Molecular characterization of the S-adenosyl-L-methionine:  
 CC 3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase involved in  
 CC isoprenolide alkaloid biosynthesis in Coptis japonica.";  
 CC RL J. Biol. Chem. 275:23398-23405(2000).  
 CC CC -1- FUNCTION: CATALYZES THE TRANSFER OF THE METHYL GROUP TO THE 4'-  
 CC HYDROXYL GROUP OF 3'-HYDROXY-N-METHYLCOCCLAURINE TO FORM  
 CC RETICULINE.  
 CC CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 3'-hydroxy-N-methyl-  
 CC (S)-coclaurine = S-adenosyl-L-homocysteine + (S)-reticuline.  
 CC CC -1- PATHWAY: CARRIES OUT A STEP IN THE CONVERSION OF 3'-HYDROXY-N-  
 CC METHYLCOCCLAURINE TO RETICULINE, AN IMPORTANT INTERMEDIATE IN  
 CC SYNTHESIZING ISOQUINOLINE ALKALOIDS.  
 CC CC -1- SUBUNIT: HOMODIMER.  
 CC CC -1- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: D29812; BAB08005.1; -  
 DR InterPro: IPR001601; Meth-transf.  
 DR InterPro: IPR001077; Methyltransf-2.  
 DR Pfam: PF00891; Methyltransf.2; 1.  
 DR K M Transferrase; Methyltransferase.  
 DR SEQUENCE 350 AA; 38775 MW; 547835EBCDEF9182 CRC64;  
 CC -----  
 CC Query Match 42.6%; Score 43; DB 1; Length 350;  
 CC Best Local Similarity 53.3%; Pred. No. 18;  
 CC Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 CC -----  
 CC Y 1 LEEIANSKDPNM 15  
 CC :||:|||||:  
 CC Db 219 DLPHVANSYDLPMI 233  
 CC -----  
 CC RESULT 15  
 CC ID RPOB\_CAMJE STANDARD; PRT; 1378 AA.  
 CC AC 046124; 09PI31;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase  
 CC beta chain) (RNA polymerase beta subunit).  
 CC GN RPOB OR C10478.  
 CC OS Campylobacter jejuni.  
 CC OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 CC OC Campylobacter.  
 CC OX NCBI\_Taxid=197;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=NCTC 11168;  
 CC RX MEDLINE=20150912; PubMed=10688204;  
 CC RA Barkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 CC Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 CC Jørgensen K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,  
 CC Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 CC Whitehead S., Barrett B.G.;  
 CC RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 CC RT reveals hypervariable sequences.";  
 CC RL Nature 403:665-668(2000).  
 CC CC [2]  
 CC RN [2]  
 CC RP SEQUENCE OF 338-1031 FROM N.A.  
 CC RX MEDLINE=96084944; PubMed=7489896;  
 CC RA Bustamante V.H., Puente J.L., Sanchez-Lopez F., Bobadilla M.,  
 CC RA Calva E.;  
 CC RT "Identification of Campylobacter jejuni and C. coli using the rpoB  
 CC RT gene and a cryptic DNA fragment from C. jejuni.";  
 CC RL Gene 165:1-8(1995).  
 CC CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC CC SUBSTRATES.  
 CC CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC (RNA)(N).  
 CC CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC CC BETA CHAIN.  
 CC CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AL139075; CAB75116.1; -  
 CC EMBL: X77304; CAAS4509.1; -  
 CC InterPro: IPR001572; RNA\_pol\_B.  
 CC Pfam: PF00362; RNA\_pol\_B; 1.  
 CC -----

DR PROSITE: PS01166; RNA\_POL\_BETA; 1.  
KW Transferrase; Transcription; DNA-directed RNA polymerase;  
Complete proteome.  
FT CONFLICT 338 347 NDLANGVDA -> MTWLMALMOP (IN REF. 2).  
FT CONFLICT 558 558 A -> R (IN REF. 2).  
FT CONFLICT 671 671 C -> S (IN REF. 2).  
FT CONFLICT 691 691 A -> R (IN REF. 2).  
SQ SEQUENCE 1378 AA; 155915 MW; AB7467C305028EB5 CRC64;

Query Match 41.18; Score 41.5; DB 1; Length 1378;  
Best Local Similarity 47.48; Pred. No. 1.3e+02;  
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

OY 2 LEEIANSKDPFNMSETDL 20  
Db 858 IEEI---TKDIPNKEDV 873

RESULT 16  
ID REAF\_HAEIN STANDARD; PRT; 346 AA.  
AC P45042;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ADP-heptose--LPS heptosyltransferase II (EC 2.-.-.-).  
GN REAF OR H1105.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=127;  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fline L.D., Friedman J.L., Geoghagen N.S.M.,  
RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -1- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 9.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U33790; AAC22760.1; -.  
DR TIGR, H11105; -.  
DR InterPro: IPR002201; Glyco\_transf\_9.  
DR Pfam: PF01075; Glyco\_transf\_9; 1.  
KW Lipopolysaccharide biosynthesis; Transferrase; Glycosyltransferase;  
Complete proteome.  
SQ SEQUENCE 346 AA; 38847 MW; 14D0E11C6773FCA9 CRC64;

Query Match 40.68; Score 41; DB 1; Length 346;  
Best Local Similarity 50.08; Pred. No. 36;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 LEEIANSKDPFNM 15

Db 116 LNDLRANKDYPMM 129

RESULT 17  
ID PT16\_HUMAN STANDARD; PRT; 376 AA.  
AC P35237;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Placental thrombin inhibitor (Cytoplasmic antiprotease) (CAP)  
DE (Protease inhibitor 6).  
GN SERPINB6 OR P16 OR PTI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94022386; PubMed=8415716.  
RA Coughlin P., Sun J., Cerruti L., Salem H.H., Bird P.;  
RT "Cloning and molecular characterization of a human intracellular  
RT serine proteinase inhibitor."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:9417-9421(1993).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-73 AND 144-149.  
RC TISSUE=Placenta;  
RX MEDLINE=94183847; PubMed=8136380.  
RA Morgenstern K.A., Sprecher C.A., Holth L., Foster D., Grant F.J.,  
RA Ching A., Kistiel W.;  
RT "Complementary DNA cloning and kinetic characterization of a novel  
RT intracellular serine proteinase inhibitor: mechanism of action with  
RT trypsin and factor Xa as model proteinases."  
RL Biochemistry 33:3432-3441(1994).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE. ALSO FOUND  
CC IN PLACENTA, CARDIAC MUSCLE, LUNG, LIVER, KIDNEY AND PANCREAS.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z22658; CA80373.1; -.  
DR EMBL: S69272; AAB30320.1; -.  
DR PIR: S35750; S35750.  
DR PIR: A48681; A48681.  
DR HSP: P05120; IBY7.  
DR MIM: 173321; -.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin; 1.  
DR SMART: SM00093; SERPIN; 1.  
DR PROSITE: PS00284; SERPIN; 1.  
KW Serpin; Serine protease inhibitor.  
FT ACT\_SITE 341 342 REACTIVE BOND.  
FT CONFLICT 175 175 G -> E (IN REF. 2).  
FT CONFLICT 362 362 R -> S (IN REF. 2).  
SQ SEQUENCE 376 AA; 42587 MW; 2BF17B347F91D7E9 CRC64;

Query Match 40.68; Score 41; DB 1; Length 376;  
Best Local Similarity 33.38; Pred. No. 39;  
Matches 10; Conservative 3; Mismatches 7; Indels 10; Gaps 1;

OY 1 LDEIANSKDPFNMSETDL 20  
Db 281 DMESVLRNLCMTDAFEKADFGSGMSQDTL 310

RESULT 18  
CLOS\_CLOHI STANDARD; PRT; 526 AA.  
ID CLOS\_CLOHI  
AC P09870: P09869;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Alpha-clostripain precursor (EC 3.4.22.8) (Clostridiopeptidase B).  
OS CLOSI.  
GN Clostridium histolyticum.  
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1498;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93341452; PubMed=8341259;  
RA Dargatz H., Dieffenhal T., Witte V., Reipen G., von Wettstein D.;  
RT "The heterodimeric protease clostripain from Clostridium histolyticum  
is encoded by a single gene."  
RT Mol. Gen. Genet. 240:140-145(1993).  
RN [2]  
RP SEQUENCE OF 51-181.  
RX MEDLINE=85076641; PubMed=6391922;  
RA Gillies A.M., Lecroisey A., Kell B.;  
RT "Primary structure of alpha-clostripain light chain."  
RT Eur. J. Biochem. 145:469-476(1984).  
RN [3]  
RP PRELIMINARY SEQUENCE OF 51-73 AND 191-232.  
RX MEDLINE=83131688; PubMed=6337850;  
RA Gillies A.M., de Wolf A., Kell B.;  
RT "Amino-acid sequences of the active-site sulfhydryl peptide and other  
thiol peptides from the cysteine proteinase alpha-clostripain."  
RT Eur. J. Biochem. 130:473-479(1983).  
CC -1- FUNCTION: CYSTEINE ENDOPEPTIDASE WITH STRICT SPECIFICITY.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, including  
Arg-I-Pro bond, but not Lys-I-Xaa.  
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND AN HEAVY CHAIN HELD  
TOGETHER BY STRONG NONCOVALENT FORCES RATHER THAN BY  
INTRAMOLECULAR DISULFIDE BRIDGES.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C11.  
CC -1- DATABASE: NAME-Worthington enzyme manual;  
WWW-http://www.worthington-biochem.com/manual/C/CP.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: X63673; CAA45212.1; -  
DR PIR: A29174; A29174.  
DR PIR: A29175; A29175.  
DR PIR: B29175; B29175.  
DR PIR: S35190; S35190.  
DR MEROPS: C11.001; -  
KW Hydrolase; Thiol protease; zymogen; Signal.  
FT SIGNAL 1 27  
FT PROPEP 28 50  
FT CHAIN 51 181  
FT PEPTIDE 182 190  
FT CHAIN 191 526  
FT ACT SITE 231 231  
FT CONFLICT 127 127  
FT CONFLICT 176 179  
FT CONFLICT 197 197  
FT CONFLICT 213 213  
FT CONFLICT 216 216  
FT CONFLICT 232 232  
SQ SEQUENCE 526 AA; 59733 MW; 1151372FF6C95BE7 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 526;  
Best Local Similarity 43.8%; Pred. No. 56;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
OY 1 DIEBIANSKDFPMS 16  
Db 75 DIEEMKTGYKSPMLN 90  
RESULT 19  
ID UDB4\_HUMAN STANDARD; PRT; 528 AA.  
AC P06133; P36538; O60731; O60867; O75614;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE UDP-glucuronosyltransferase 2B4 precursor, microsomal (EC 2.4.1.17)  
DE (UDPgt) (Hydroxycholeic acid) (HUG25) (UDPgtH-1).  
GN UGT2B4 OR UGT2B11.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87241362; PubMed=3109396;  
RA Jackson M.R., McCarthy L.R., Harding D., Wilson S., Coughtrie M.W.H.,  
RT "Cloning of a human liver microsomal UDP-glucuronosyltransferase  
cDNA."  
RT Biochem. J. 242:581-588(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=93326164; PubMed=8333863;  
RA Jin C.-J., Miners J.O., Lillywhite K.J., McKenzie P.I.;  
RT "CDNA cloning and expression of two new members of the human liver  
UDP-glucuronosyltransferase 2B subfamily."  
RT Biochem. Biophys. Res. Commun. 194:496-503(1993).  
RN [3]  
RP SEQUENCE FROM N.A., VARIANT GLU-458, AND CHARACTERIZATION.  
RX PubMed=10376768;  
RA Levesque E., Beaulieu M., Hum D.W., Belanger A.;  
RT "Characterization and substrate specificity of UGT2B4 (E458): a  
UDP-glucuronosyltransferase encoded by a polymorphic gene."  
RT Pharmacogenetics 9:207-216(1999).  
RN [4]  
RP SEQUENCE FROM N.A., AND VARIANTS LEU-109 AND LEU-396.  
RA McKenzie P.I.;  
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: UDPgt's are of major importance in the conjugation and  
CC subsequent elimination of potentially toxic xenobiotics and  
CC endogenous compounds. This isozyme is active on polyhydroxylated  
CC estrogens (such as estradiol, 4-hydroxyestrone and 2-hydroxyestradiol)  
CC and xenobiotics (such as 4-methylumbelliferone, 1-naphthol, 4-  
CC nitrophenol, 2-aminophenol, 4-hydroxybiphenyl and menthol). It is  
CC capable of 6 alpha-hydroxyglucuronidation of hydroxycholeic acid.  
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor  
CC beta-D-glucuronoside.  
CC -1- SUBCELLULAR LOCATION: Microsomal.  
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
CC -1- CAUTION: REF. 2 THOUGHT THAT THIS WAS A NEW FORM (UGT2B11). THE  
CC NAME UGT2B11 HAS NOW BEEN REUSED FOR ANOTHER HUMAN ENZYME.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to [license@lsb.slb.ch](mailto:license@lsb.slb.ch).  
-----  
DR EMBL: Y00317; CAA68415.1; -  
DR EMBL: AF064200; AAC95002.1; -  
DR EMBL: A1005162; CAA06396.1; -  
DR EMBL: AF081793; AAC32272.1; -  
DR PIR: A27878; A27878.  
DR PIR: JN0619; JN0619.  
DR MIM: 600067; -  
DR InterPro: IPR002213; UDPGT.  
DR Pfam: PF00201; UDPGT. 1.  
DR PROSITE: PS00375; UDPGT. 1.  
KW transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;  
KW Multigene family; Microsome; Polymorphism.  
FT SIGNAL 1 23 UDP-GLUCONOSYLTRANSFERASE 2B4.  
FT CHAIN 24 528  
FT TRANSMEM 493 509  
FT CARBOHYD 315 315  
FT VARIANT 109 109  
FT VARIANT 396 396  
FT VARIANT 458 458  
FT VARIANT 171 172  
FT CONFLICT 291 293  
FT CONFLICT 382 387  
FT SEQUENCE 528 AA; 60512 MW; 6B45E6769971A078 CRC64;  
-----  
Query Match 40.6%; Score 41; DB 1; Length 528;  
Best Local Similarity 66.7%; Pred. No. 57;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
-----  
QY 6 AANSKDFPNMSETDL 20  
DB 411 AAVSLDFHTMSSTD 425  
-----  
RESULT 20  
HLVB\_VTBCH STANDARD; PRT; 548 AA.  
ID HLVB\_VTBCH  
AC P15492; Q9KM08;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemolysin secretion protein precursor.  
GN HLVB OR VCA0220.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID:666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR 017 / SEROTYPE O1;  
MEDLINE=90286918; PubMed=2162464;  
RA Alm R.A., Manning P.A.;  
RT "Characterization of the hlyB gene and its role in the production of  
the El Tor haemolysin of Vibrio cholerae O1.";  
RL Mol. Microbiol. 4:413-425(1990).  
RN [2]  
RP REVISIONS.  
RC STRAIN-EL TOR 017 / SEROTYPE O1;  
Manning P.A.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
Emolajeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae.";  
RL Nature 406:477-483(2000).  
CC -I- FUNCTION: HLYB PROTEIN MAY FORM A PORE THROUGH WHICH THE  
HEMOLYSIN CAN BE EXPORTED.  
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.lsb.slb.ch/announce/>  
CC or send an email to [license@lsb.slb.ch](mailto:license@lsb.slb.ch)).  
-----  
DR EMBL: Y00557; CAA68368.1; -  
DR EMBL: AE004362; AAF96132.1; -  
DR PIR: S15910; S15910.  
DR HSSP: P02942; 10U7.  
DR TIGR: VCA0220; -  
DR InterPro: IPR004089; Chemotaxis\_transducer.  
DR InterPro: IPR003660; HAM.  
DR InterPro: IPR004090; Me\_Chemotaxis.  
DR Pfam: PF00672; HAM; 1.  
DR Pfam: PF00015; MCPsignal; 1.  
DR PRINTS: PR00260; CHEMTRNSDCCR.  
DR SMART: SM00304; HAM; 1.  
DR SMART: SM00283; MA; 1.  
KW Hemolysis; Transport; Transmembrane; Signal; Complete proteome.  
FT SIGNAL 1 17  
FT CHAIN 18 548  
FT TRANSMEM 23 52  
FT TRANSMEM 198 220  
FT TRANSMEM 366 382  
FT TRANSMEM 22 22  
FT CONFLICT 103 103  
FT CONFLICT 181 182  
FT CONFLICT 330 330  
FT CONFLICT 334 335  
FT SEQUENCE 548 AA; 60034 MW; E89211DD8038A2P4 CRC64;  
-----  
Query Match 40.6%; Score 41; DB 1; Length 548;  
Best Local Similarity 41.2%; Pred. No. 59;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
-----  
QY 2 LEEIANSKDFPNMSET 18  
DB 270 IEEVAENSKELATMASS 286  
-----  
RESULT 21  
RUAP\_SOYBN STANDARD; PRT; 283 AA.  
ID RUAP\_SOYBN  
AC P39657;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE RUBISCO-associated protein.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=leaf;  
RX MEDLINE=95067276; PubMed=7972504;  
RA Staswick P.E., Crafts-Brandner S.J., Salvucci M.E.;  
RT "cDNA sequence for the ribulose 1,5 biphosphate  
carboxylase/oxygenase complex protein. A protein that accumulates in

```

RT soybean leaves in response to fruit removal.",
CC Plant Physiol. 105:1445-1446(1994).
CC -1- SUBUNIT: FORMS PART OF THE RUBISCO COMPLEX.
CC -1- TISSUE SPECIFICITY: LEAVES.
CC -1- INDUCTION: BY FRUIT REMOVAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L28804; AAA34007.1;
DR InterPro: IPR000677; 2S_Globulin.
DR Pfam: PF02220; Nardonnin; 1.
SQ SEQUENCE 283 AA; 31258 MW; 876A669B1E532560 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 283;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 EBIANSKDPN 14
DB 128 EDIANEADFN 139

RESULT 22
LDD_HAEIN STANDARD; PRT; 381 AA.
ID LDD_HAEIN
AC P46454;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE L-lactate dehydrogenase (Cytochrome) (EC 1.1.2.3).
GN LLD OR H11739.1.
OS Haemophilus Influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Colton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RA Koonin E.V., Rudd K.E.;
RT Submitted (SEP-1995) to the SWISS-PROT data bank.
CC -1- CATALYTIC ACTIVITY: (S)-lactate + 2 ferriocytochrome c - pyruvate +
CC 2 ferriocytochrome c.
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FMN-DEPENDENT ALPHA-HYDROXY ACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32847; AAC23385.1;
DR HSSP: P00175; ILDC.
DR TIGR: H11739.1;
DR InterPro: IPR003009; FMN_enzyme.
DR InterPro: IPR000262; FMN_hydroxy_acid_dh.
DR Pfam: PF01070; FMN_dh; 1.
DR PROSITE: PS00557; FMN-HYDROXY ACID_DH; 1.
KW Oxidoreductase; Flavoprotein; FMN: Complete proteome.
FT ACT_SITE 24 24 SUBSTRATE BINDING (BY SIMILARITY).
FT ACT_SITE 129 129 SUBSTRATE BINDING (BY SIMILARITY).
FT ACT_SITE 275 275 REMOVES THE SUBSTRATE ALPHA-PROTON AS THE
FT ACT_SITE FIRST STEP IN CATALYSIS (BY SIMILARITY).
FT ACT_SITE 278 278 SUBSTRATE BINDING (BY SIMILARITY).
SQ SEQUENCE 381 AA; 41965 MW; E1B6280A1A6482D CRC64;

Query Match 39.6%; Score 40; DB 1; Length 381;
Best Local Similarity 47.4%; Pred. No. 57;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DLEIANSKDPNSTD 19
DB 41 DLENIALRQRLKDMSELD 59

RESULT 23
LUTP_HUMAN STANDARD; PRT; 547 AA.
ID LUTP_HUMAN
AC P22307; Q15432; Q99430;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1991 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Nonspecific lipid-transfer protein, mitochondrial precursor (NSL-TP)
DE (Sterol carrier protein 2) (SCP-2) (Sterol carrier protein X) (SCP-X)
DE (SCPX).
GN SCP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95213031; PubMed=7698762;
RA Ohba T., Rennert H., Pfeifer S.M., He Z., Yamamoto R., Holt J.A.,
RA Billheimer J.T., Strauss J.F. III;
RT "The structure of the human sterol carrier protein X/sterol carrier
RT protein 2 gene (SCP2)."
RL Genomics 24:370-374(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92029618; PubMed=1718316;
RA He Z., Yamamoto R., Furtz E.E., Schantz L.J., Naylor S.L., George H.,
RA Billheimer J.T., Strauss J.F. III;
RT "cDNAs encoding members of a family of proteins related to human
RT sterol carrier protein 2 and assignment of the gene to human
RT chromosome 1 p21-pter."
RL DNA Cell Biol. 10:559-569(1991).
RN [3]
RP SEQUENCE OF 405-547 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91110550; PubMed=1703300;
RA Yamamoto R., Kallen C.B., Babalola G.O., Rennert H., Billheimer J.T.,
RA Strauss J.F. III;
RT "Cloning and expression of a cDNA encoding human sterol carrier
RT protein 2."
RL Proc. Natl. Acad. Sci. U.S.A. 88:463-467(1991).
RN [4]
RP SEQUENCE OF 405-547 FROM N.A.

```



RC TISSUE-LIVER;  
 RA MEDLINE=93131254; PubMed=1483685;  
 RX Yamamoto R.;  
 RT "Localization of human sterol carrier protein 2 gene and cDNA  
 expression in COS-7 cell.";  
 RL Hokkaido Igaku Zasshi 67:839-848(1992).  
 RN [5]  
 RP STRUCTURE BY NMR OF SCP2.  
 RX MEDLINE=94063072; PubMed=8243660;  
 RA Szyperski T., Scheek S., Johansson J., Assmann G., Seedorf U.,  
 RW Wuelrich K.;  
 RT "NMR determination of the secondary structure and the  
 three-dimensional polypeptide backbone fold of the human sterol  
 carrier protein 2.";  
 RL FEBS Lett. 335:18-26(1993).  
 CC -1- FUNCTION: MEDIATES IN VITRO THE TRANSFER OF ALL COMMON  
 PHOSPHOLIPIDS, CHOLESTEROL AND GANGLIOSIDS BETWEEN MEMBRANES. MAY  
 PLAY A ROLE IN REGULATING STEROIDOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC IN THE LIVER AND ALSO ASSOCIATED  
 WITH MITOCHONDRIA ESPECIALLY IN STEROIDOGENIC TISSUES. SCP-X  
 CC RESIDES IN THE PEROXISOME.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENTIAL INITIATION OF SCP2 GENE  
 TRANSCRIPT LEADS TO THE EXPRESSION OF PRE-SCP-2 AND SCP-X FROM A  
 SINGLE GENE.  
 CC -1- TISSUE SPECIFICITY: LIVER, FIBROBLASTS, AND PLACENTA.  
 CC -1- DISEASE: SCP2 IS PRESENT IN LOW LEVELS IN SUBJECTS WITH ZELLWEGER  
 SYNDROME (CEREBRO-HEPATIC-RENAL SYNDROME), WHOSE CELLS ARE  
 DEFICIENT IN PEROXISOMES AND WHO HAVE AN ASSOCIATED IMPAIRMENT IN  
 PLASMA/GEN AND BILE ACID SYNTHESIS AND CATABOLISM OF PHYTANIC  
 ACID AND VERY-LONG-CHAIN FATTY ACIDS.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE THIOLEASE  
 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U11313; AAB41286.1; -  
 DR EMBL: U11297; AAB41286.1; JOINED.  
 DR EMBL: U11299; AAB41286.1; JOINED.  
 DR EMBL: U11301; AAB41286.1; JOINED.  
 DR EMBL: U11302; AAB41286.1; JOINED.  
 DR EMBL: U11303; AAB41286.1; JOINED.  
 DR EMBL: U11304; AAB41286.1; JOINED.  
 DR EMBL: U11305; AAB41286.1; JOINED.  
 DR EMBL: U11306; AAB41286.1; JOINED.  
 DR EMBL: U11307; AAB41286.1; JOINED.  
 DR EMBL: U11308; AAB41286.1; JOINED.  
 DR EMBL: U11309; AAB41286.1; JOINED.  
 DR EMBL: U11310; AAB41286.1; JOINED.  
 DR EMBL: U11311; AAB41286.1; JOINED.  
 DR EMBL: U11312; AAB41286.1; JOINED.  
 DR EMBL: M75883; AAA03557.1; -  
 DR EMBL: M75884; AAA03558.1; ALT\_INT.  
 DR EMBL: M55421; AAA03559.1; ALT\_SEO.  
 DR EMBL: S52450; AAB24921.1; -  
 DR PIR: A39010; A39010.  
 DR HSSP: P27796; 1PXT.  
 DR MIM: 184755; -  
 DR InterPro: IPR003033; SCP2.  
 DR InterPro: IPR002155; Thiolease.  
 DR Pfam: PF02036; SCP2; 1.  
 DR Pfam: PF02108; Thiolease; 1.  
 DR Pfam: PF02803; Thiolease\_C; 1.  
 DR PROSITE: PS00098; THIOLEASE\_1; 1.  
 DR PROSITE: PS00737; THIOLEASE\_2; 1.  
 DR PROSITE: PS00099; THIOLEASE\_3; FALSE\_NEG.  
 DR PROSITE: PS00342; MICROBODIES\_CTER; 1.

KW Lipid-binding; Transport; Mitochondrion; Peroxisome; Transit peptide;  
 KM Alternative Initiation.  
 FT CHAIN 1 547 SCP-X.  
 FT TRANSIT 405 424 MITOCHONDRION (BY SIMILARITY).  
 FT CHAIN 425 547 NONSPECIFIC LIPID-TRANSFER PROTEIN.  
 FT INIT\_MET 405 405 FOR NSL-TP.  
 FT ACT\_SITE 94 94 SUBSTRATE BINDING (BY SIMILARITY).  
 FT SITE 494 494 ESSENTIAL FOR TRANSPORT OF LIPIDS.  
 FT SITE 545 547 MICROBODY TARGETING SIGNAL (POTENTIAL).  
 FT CONFLICT 10 10 T -> A (IN REF. 1).  
 FT CONFLICT 393 393 G -> D (IN REF. 1).  
 FT CONFLICT 472 472 A -> D (IN REF. 3 AND 4).  
 FT CONFLICT 482 482 K -> Q (IN REF. 3 AND 4).  
 FT CONFLICT 501 501 D -> A (IN REF. 3 AND 4).  
 FT CONFLICT 522 522 K -> P (IN REF. 3 AND 4).  
 SO SEQUENCE 547 AA; 58993 MW; 29F7551465C7143A CRC64;  
 Query Match 39.6%; Score 40; DB 1; Length 547;  
 Best Local Similarity 50.0%; Pred. No. 84;  
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 Oy 6 AANSKDPNMS 17  
 Db 29 AENSNDYDPLAE 40  
 ID NSFH\_CAEEL STANDARD; PRT; 824 AA.  
 AC 094392;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Vesicular-fusion protein NSF homolog (N-ethylmaleimide-sensitive  
 DE fusion protein) (NEM-sensitive fusion protein).  
 GN H15N14.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodidae; Caenorhabditis.  
 OC NCBI\_Taxid=6239;  
 OX  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Wilkinson J., McMurray A.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RP REVISIONS.  
 RA Durbin R.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR VESICLE-MEDIATED TRANSPORT. CATALYZES THE  
 CC FUSION OF TRANSPORT VESICLES WITHIN THE GOLGI CISTERNAE. IS ALSO  
 CC REQUIRED FOR TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO THE GOLGI  
 CC STACK. SEEM TO FUNCTION AS A FUSION PROTEIN REQUIRED FOR THE  
 CC DELIVERY OF CARGO PROTEINS TO ALL COMPARTMENTS OF THE GOLGI STACK  
 CC INDEPENDENT OF VESICLE ORIGIN (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOHETEROMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z79698; CAB01976.1; -  
 DR EMBL: Z96100; CAB01976.1; JOINED.  
 DR EMBL: Z96100; CAB09531.1; -  
 DR EMBL: Z79698; CAB09531.1; JOINED.  
 DR HSSP: P18708; NSF.

DR WormRep: H15N14.1: CE19925.  
DR InterPro: IPR003559; AAA.  
DR InterPro: IPR003960; AAA\_sub.  
DR InterPro: IPR003959; AAA\_subfam.  
DR InterPro: IPR003338; VAT\_Nu.  
DR InterPro: IPR004201; cdc48\_2.  
DR Pfam: PF000004; AAA\_1.  
DR Pfam: PF02933; cdc48\_2; 1.  
DR Pfam: PF02359; cdc48\_N; 1.  
DR SMART: SM00382; AAA; 2.  
DR PROSITE: PS00674; AAA; 1.  
KM Transport: Protein transport: Endoplasmic reticulum; Golgi stack;  
KM ATP-binding; Repeat.  
FT NP\_BIND 337 344 ATP (POTENTIAL).  
FT NP\_BIND 620 627 ATP (POTENTIAL).  
SQ SEQUENCE 824 AA; 91334 MW; 67232C5115B0B3A CRC64;

Query Match 39.6%; Score 40; DB 1; Length 824;  
Best Local Similarity 42.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 LEEIANSKDFPNMSETDL 20  
DB 547 IEKLAINGSDPYALENDI 565

RESULT 25  
EAL3\_HUMAN STANDARD; PRT; 1087 AA.  
AC 09Y2J2; Q9BRP5; 095713;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DE Band 4.1-like protein 3 (4.1B) (Differentially expressed in  
adenocarcinoma of the lung protein 1) (DAL-1).  
OS EPR41L3 OR DAL1 OR KIA0987.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE=Brain;  
RX MEDLINE=99246063; PubMed=10231032;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:63-70(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM C).  
RC TISSUE=Lung;  
RX MEDLINE=99107198; PubMed=9892180;  
RA Tran Y.K., Boegler O., Gorse K.M., Wieland I., Green M.R.,  
RA Newsham I.F.;  
RT "A novel member of the NF2/ERM/4.1 superfamily with growth suppressing  
RT properties in lung cancer.";  
RL Cancer Res. 59:35-43(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Lung;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Critical growth regulator in the pathogenesis of  
CC meningiomas.  
CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; A (shown here), B and  
CC C; are produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Expressed at high levels in brain, with lower  
CC levels in kidney, intestine, and testis.  
CC -1- CAUTION: Ref.2 sequence differs from that shown due to  
CC frameshifts in position 29 and 59.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AB023204; BAA76831.1; ALT\_INT.  
DR EMBL: AF069072; AAC79806.1; ALT\_FRAME.  
DR EMBL: BC006141; AAH06141.1; -.  
DR MIM: 603331; -.  
DR InterPro: IPR000299; Band\_4.1.  
DR Pfam: PF00373; Band\_41; 1.  
DR PRINTS: PR00935; BAND41.  
DR SMART: SM00295; BA1; 1.  
DR PROSITE: PS00660; BAND\_41\_1; 1.  
DR PROSITE: PS00661; BAND\_41\_2; 1.  
DR PROSITE: PS00557; BAND\_41\_3; 1.  
KM Structural protein: Cytoskeleton; Alternative splicing.  
FT DOMAIN 107 321 BAND 4.1-LIKE.  
FT DOMAIN 388 513 HYDROPHILIC.  
FT DOMAIN 514 860 SPECTRIN--ACTIN-BINDING (POTENTIAL).  
FT DOMAIN 861 1083 CARBOXYL-TERMINAL (CTD).  
FT VARSPPLIC 446 446 G-> GASVENEHEIYKDSMSAA (IN ISOFORM B  
FT AND ISOFORM C).  
FT VARSPPLIC 503 689 MISSING (IN ISOFORM B AND ISOFORM C).  
FT VARSPPLIC 708 719 MISSING (IN ISOFORM B AND ISOFORM C).  
FT VARSPPLIC 784 824 MISSING (IN ISOFORM B AND ISOFORM C).  
FT VARSPPLIC 835 1087 MISSING (IN ISOFORM C).  
FT CONFLICT 32 32 MISSING (IN REF. 2).  
FT CONFLICT 498 498 R-> Q (IN REF. 2).  
SQ SEQUENCE 1087 AA; 120677 MW; 0A33CA4A43F12620 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 1087;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 12 FPNMSETDL 20  
DB 608 FPNMSETNL 616

Search completed: June 16, 2002, 00:10:12  
Job time: 660 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:54 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-45

Perfect score: 101

Sequence: 1 DLEETRANSKDFPNMSETDL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

## Database :

SPTREMBL\_19:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_vivirus:\*  
16: sp.\_bacteriophage:\*  
17: sp.\_archaeoprotein:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	251	12 09ENR1	09enr1 hepatitis a
2	101	100.0	251	12 09ENR9	09enr9 hepatitis a
3	101	100.0	251	12 09ENR6	09enr6 hepatitis a
4	101	100.0	251	12 09ENR5	09enr5 hepatitis a
5	101	100.0	251	12 09ENR4	09enr4 hepatitis a
6	101	100.0	251	12 09ENR1	09enr1 hepatitis a
7	101	100.0	251	12 09ENR9	09enr9 hepatitis a
8	101	100.0	251	12 09ENR7	09enr7 hepatitis a
9	101	100.0	251	12 09ENR5	09enr5 hepatitis a
10	101	100.0	251	12 09ENR2	09enr2 hepatitis a
11	101	100.0	251	12 09ENR1	09enr1 hepatitis a
12	101	100.0	251	12 09ENR7	09enr7 hepatitis a
13	101	100.0	251	12 09ENR6	09enr6 hepatitis a
14	101	100.0	251	12 09ENR5	09enr5 hepatitis a
15	101	100.0	251	12 09ENR4	09enr4 hepatitis a
16	101	100.0	251	12 09ENR2	09enr2 hepatitis a

## ALIGNMENTS

17	101	100.0	1124	12 084780	084780 hepatitis a
18	101	100.0	1161	12 005794	005794 hepatitis a
19	101	100.0	2216	12 09WMA2	09WMA2 hepatitis a
20	101	100.0	2218	12 067824	067824 hepatitis a
21	101	100.0	2225	12 09D32	09D32 hepatitis a
22	101	100.0	2227	12 067825	067825 hepatitis a
23	101	100.0	2227	12 067826	067826 hepatitis a
24	101	100.0	2227	12 09WMA4	09WMA4 hepatitis a
25	101	100.0	2227	12 09WMA3	09WMA3 hepatitis a
26	101	100.0	2227	12 09WMA1	09WMA1 hepatitis a
27	101	100.0	2227	12 09WMA0	09WMA0 hepatitis a
28	101	100.0	2227	12 091FH5	091FH5 hepatitis a
29	95	94.1	251	12 09ENR8	09enr8 hepatitis a
30	95	94.1	251	12 09ENR7	09enr7 hepatitis a
31	95	94.1	251	12 09ENR3	09enr3 hepatitis a
32	95	94.1	251	12 09ENR0	09enr0 hepatitis a
33	95	94.1	251	12 09ENR8	09enr8 hepatitis a
34	95	94.1	251	12 09ENR6	09enr6 hepatitis a
35	95	94.1	251	12 09ENR4	09enr4 hepatitis a
36	95	94.1	251	12 09ENR3	09enr3 hepatitis a
37	95	94.1	251	12 09ENR0	09enr0 hepatitis a
38	95	94.1	251	12 09ENR9	09enr9 hepatitis a
39	95	94.1	251	12 09ENR8	09enr8 hepatitis a
40	95	94.1	251	12 09ENR3	09enr3 hepatitis a
41	95	94.1	2218	12 067817	067817 hepatitis a
42	94	93.1	184	12 087092	087092 simian hepa
43	91	90.1	251	12 09ENR2	09enr2 hepatitis a
44	91	90.1	2227	12 09WMA9	09WMA9 hepatitis a
45	90	89.1	251	12 09ENR0	09enr0 hepatitis a
46	90	89.1	736	16 098Q68	098Q68 mycoplasma
47	49	48.5	400	5 016961	016961 caenorhabdi
48	46	45.5	424	5 09VW80	09VW80 drosophila
49	45	45.0	930	10 09ZVD5	09ZVD5 arabidopsis
50	45	45.0	1621	5 001900	001900 caenorhabdi
51	44	44.1	1712	5 09V839	09V839 drosophila
52	44	43.6	1046	3 007653	007653 saccharomyc
53	43	43.1	225	4 09NR10	09NR10 homo sapien
54	43	43.1	299	4 09H291	09H291 homo sapien
55	43	43.5	5071	5 P91905	P91905 caenorhabdi
56	43	43.1	5107	5 094279	094279 caenorhabdi
57	43	42.6	152	16 025423	025423 helicobacte
58	43	42.6	152	16 09ZLC2	09ZLC2 human immun
59	43	42.6	210	15 09JDD9	09JDD9 human immun
60	43	42.6	210	15 09JDC4	09JDC4 human immun

RESULT 1  
ID 09ENR1 PRELIMINARY; PRT: 251 AA.

AC 09ENR1; 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE POLYPROTEIN (FRAGMENT).

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RC STRAIN=AL;

RA Fujinawa K.;

RT "hepatitis A virus."

DR EMBL: AB047652; BABI2160.1; to the EMBL/GenBank/DBJ databases.

FT NON\_TER 1 251

SO SEQUENCE 251 AA; 28749 MW; 58A520D873893445 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 2  
O9EN09 PRELIMINARY; PRT; 251 AA.  
AC O9EN09;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A159;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047654; BAB12162.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 3  
O9EN06 PRELIMINARY; PRT; 251 AA.  
AC O9EN06;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A162;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047657; BAB12165.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28639 MW; 56AA86DF60F6FD3B CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 4  
O9EN05 PRELIMINARY; PRT; 251 AA.  
AC O9EN05;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A20;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047658; BAB12166.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 5  
O9EN04 PRELIMINARY; PRT; 251 AA.  
AC O9EN04;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A201;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047659; BAB12167.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 6  
O9EN01 PRELIMINARY; PRT; 251 AA.  
AC O9EN01;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)

01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A206;  
RA Fujiwara K.;  
RT "hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047662; BAB12170.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28699 MW; 8EADAEE7E2754C37 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20  
Db 95 DLEETIANSKDFPNMSETDL 114  
|||||

RESULT 7  
Q9ENP9 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP9;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A303;  
RA Fujiwara K.;  
RT "hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047664; BAB12172.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28752 MW; 7215A28AD2CA5C1A CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20  
Db 95 DLEETIANSKDFPNMSETDL 114  
|||||

RESULT 8  
Q9ENP7 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP7;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-A306;  
RC Fujiwara K.;  
RA "hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047666; BAB12174.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20  
Db 95 DLEETIANSKDFPNMSETDL 114  
|||||

RESULT 9  
Q9ENP5 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP5;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A407;  
RA Fujiwara K.;  
RT "hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047668; BAB12176.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28614 MW; 8334EF179C757A6D CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETIANSKDFPNMSETDL 20  
Db 95 DLEETIANSKDFPNMSETDL 114  
|||||

RESULT 10  
Q9ENP2 PRELIMINARY; PRT; 251 AA.  
AC Q9ENP2;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A5;  
RA Fujiwara K.;  
RT "hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047671; BAB12179.1; -.  
|||||

FT NON\_TER 1 1  
FT NON\_TER 251  
SQ SEQUENCE 251 AA; 28622 MW; B55C3CD146D39D02 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 11  
Q9ENP1 PRELIMINARY; PRT; 251 AA.

AC Q9ENP1; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A503;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047672; BAB12180.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28663 MW; C7EA66FBD19A1619 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 12

Q9ENN7 PRELIMINARY; PRT; 251 AA.

AC Q9ENN7; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A712;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047676; BAB12184.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28785 MW; EFE79D3A26134F18 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 13  
Q9ENN6 PRELIMINARY; PRT; 251 AA.

AC Q9ENN6; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A713;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047677; BAB12185.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114

RESULT 14  
Q9ENN5 PRELIMINARY; PRT; 251 AA.

AC Q9ENN5; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A75;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047678; BAB12186.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28658 MW; 98E8EDD0B2EDF10 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 6.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
DB 95 DLEEIAANSKDFPNMSETDL 114



```
RESULT 15
ID Q9ENN4 PRELIMINARY; PRT; 251 AA.
AC Q9ENN4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
RX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A77;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047679; BAB12187.1; -.
FT NON_TER 1
FT SEQUENCE 251 AA; 28734 MW; 38968782882F19D9 CRC64;
SQ

Query Match 100.0%; Score 101; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 6.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DLEEIAANSKDFPNMSETDL 20
Db 95 DLEEIAANSKDFPNMSETDL 114

RESULT 16
ID Q9ENN2 PRELIMINARY; PRT; 251 AA.
AC Q9ENN2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
RX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A9;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047681; BAB12189.1; -.
FT NON_TER 1
FT SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;
SQ

Query Match 100.0%; Score 101; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 6.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DLEEIAANSKDFPNMSETDL 20
Db 95 DLEEIAANSKDFPNMSETDL 114

RESULT 17
ID Q84780 PRELIMINARY; PRT; 1124 AA.
AC Q84780;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
SQ
```

```
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)
DE (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
RX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Ouchinnikov V.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,
RA Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prikhodko G.G.,
RA Bilinov V.M., Vasilchenko S.K., Sandakchiev L.S., Kusov Y.Y.,
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;
RL Dokl. Biochem. 285:379-383(1986).
DR EMBL; X04200; CAA27797.1; -.
DR EMBL; A11312; CAA00953.1; -.
KW Nonstructural protein.
FT NON_TER 1
FT SEQUENCE 1124 AA; 127026 MW; 38449E3D2ABDF8CA CRC64;
SQ

Query Match 100.0%; Score 101; DB 12; Length 1124;
Best Local Similarity 100.0%; Pred. No. 3.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DLEEIAANSKDFPNMSETDL 20
Db 888 DLEEIAANSKDFPNMSETDL 907

RESULT 18
ID Q05794 PRELIMINARY; PRT; 1161 AA.
AC Q05794; Q67800; Q67801; Q67803; Q67804; Q67805; Q67806;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (RC 2.7.7.48) (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
RX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilchenko S.K.,
RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,
RA Balayan M.S.;
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).
CC -I- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC EMBL; X15464; CAA33492.1; -.
DR InterPro; IPR000408; RCCL.
DR PROSITE; PS00626; RCCL.2; UNKNOWN.1.
KW Polypeptide; Coat protein; Core protein; RNA-directed RNA polymerase;
KW Hydroxylase; Thiol protease.
FT CHAIN 1
FT CHAIN 23
FT CHAIN 24
FT CHAIN 245
FT CHAIN 246
FT CHAIN 491
FT CHAIN 492
FT CHAIN 794
FT CHAIN 795
FT CHAIN 901
FT CHAIN 1087
FT CHAIN 1088
FT CHAIN >1161
FT NON_TER 1161
FT SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;
SQ

Query Match 100.0%; Score 101; DB 12; Length 1161;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DLEEIAANSKDFPNMSETDL 20
DB 925 DLEEIAANSKDFPNMSETDL 944

RESULT 19
Q9HMA2 PRELIMINARY: PRT: 2216 AA.
ID 09HMA2
AC Q9HMA2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH3;
RX MEDLINE=21386014; PubMed=11495028;
RA Fujisawa K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;
RT Analysis of full-length hepatitis A virus genome in sera from
RL patients with fulminant and self-limited acute type A hepatitis."
DR Hepatol.35:112-119(2001).
DR EMBL: AB020566; BAA35104.1; -.
DR MEROPS: C03.005; -.
DR InterPro: IPR004004; Calic1_pol_hel.
DR InterPro: IPR000408; RCC1.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PRINTS: PR00910; RNA_helicase; 1.
DR PROSITE: PS00626; RCC1.2; UNKNOWN.1.
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FB82 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2216;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
DB 931 DLEEIAANSKDFPNMSETDL 950

RESULT 20
Q67824 PRELIMINARY: PRT: 2218 AA.
ID 067824;
AC 067824;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE GBM/ERRK RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GBM;
RX MEDLINE=94076453; PubMed=8254770;
RA Graff J., Norman A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
RL to two cell culture adapted variants."
DR EMBL: X75214; CAA53024.1; -.
DR InterPro: IPR000408; RCC1.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PRINTS: PR00910; RNA_helicase; 1.
```

```
DR PROSITE: PS00626; RCC1.2; UNKNOWN.1.
FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEINS.
FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.
SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2218;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
DB 925 DLEEIAANSKDFPNMSETDL 944

RESULT 21
Q9DL32 PRELIMINARY: PRT: 2225 AA.
ID 09DL32
AC 09DL32;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L-A-1;
RX Wang P.F., Jiang C.L., Liu J.Y., Zhang H.Y.;
RL Submitted (OCr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF314208; AAG45423.1; -.
DR MEROPS: C03.005; -.
DR InterPro: IPR004004; Calic1_pol_hel.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PRINTS: PR00910; RNA_helicase; 1.
DR PROSITE: PS00918; CALICVIRUSNS.
SQ SEQUENCE 2225 AA; 251297 MW; EBACE41B043E59B CRC64;

Query Match 100.0%; Score 101; DB 12; Length 2225;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
DB 931 DLEEIAANSKDFPNMSETDL 950

RESULT 22
Q67825 PRELIMINARY: PRT: 2227 AA.
ID 067825;
AC 067825;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE GBM/WT RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GBM;
RX MEDLINE=94076453; PubMed=8254770;
RA Graff J., Norman A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
RL to two cell culture adapted variants."
DR EMBL: X75214; CAA53024.1; -.
DR InterPro: IPR000408; RCC1.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol. 1.
DR PRINTS: PR00910; RNA_helicase; 1.
```



Search completed: June 16, 2002, 00:08:55  
job time: 793 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:01:48 ; Search time 209.1 Seconds  
(without alignments)  
10.624 Million cell updates/sec

Title: US-09-171-432a-46  
Perfect score: 96  
Sequence: 1 KINLADRMGLSGVQEIREQ 20

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	20	AAW42929	Immunogenic Hepat
2	96	100.0	20	AAW69446	Synthetic HAV P2A
3	96	100.0	25	AAW42969	Immunogenic Hepat
4	96	100.0	25	AAW69448	Synthetic HAV P2A
5	96	100.0	2227	7 AAP60066	Sequence of viral
6	96	100.0	2227	11 AAR05697	Attenuated hepati
7	96	100.0	2227	18 AAW34074	Hepatitis A virus
8	96	100.0	2227	21 AAB18607	Amino acid sequenc
9	96	100.0	2227	21 AAB18608	Amino acid sequenc
10	96	100.0	2227	21 AAB18609	Amino acid sequenc
11	43	44.8	80	AAU41911	Protonibacterium

12	43	44.8	922	22	AAW85023	Shrimp white spot
13	42	43.8	625	22	AAU49915	Protonibacterium
14	42	43.8	1035	22	AAW60277	Saccharomyces cere
15	41.5	43.2	2014	22	AAW67395	Amino acid sequenc
16	41	42.7	213	22	AAW59304	Drosophila melanog
17	41	42.7	291	22	AAW60641	Novel human diagno
18	41	42.7	332	20	AAW37809	Protein which is a
19	41	42.7	366	21	AAW60855	Mouse CD14 protein
20	41	42.7	366	21	AAW53879	A murine CD14 prot
21	41	42.7	605	22	AAU36494	Pseudomonas aerugi
22	41	42.7	708	19	AAW69844	Amino acid sequenc
23	41	42.7	715	22	ABG12254	Novel human diagno
24	41	42.7	748	19	AAW50143	Cyanobacterial phy
25	40	41.7	92	22	ABG19753	Novel human diagno
26	40	41.7	140	22	AAW69547	C glutamicum prote
27	40	41.7	144	22	AAU57103	Protonibacterium
28	40	41.7	144	22	AAW65829	Protonibacterium
29	40	41.7	154	22	ABW29032	Peptide #1683 enco
30	40	41.7	154	22	ABW34196	Peptide #1702 enco
31	40	41.7	154	22	ABW19633	Protein #1632 enco
32	40	41.7	154	22	AAW54987	Human brain expres
33	40	41.7	154	22	AAW67372	Human bone marrow
34	40	41.7	154	22	AAW15204	Peptide #1638 enco
35	40	41.7	154	22	AAW27663	Peptide #1700 enco
36	40	41.7	154	22	AAW02946	Peptide #1628 enco
37	40	41.7	315	22	AAW1918	Human olfactory re
38	40	41.7	315	22	AAW27272	Human olfactory re
39	40	41.7	332	11	AAW03167	Amino acid sequenc
40	40	41.7	423	21	AAW01670	Yeast type II topo
41	40	41.7	440	20	AAW34626	Chlamydia pneumoni
42	40	41.7	549	22	AAW35087	Enterococcus faeca
43	40	41.7	615	22	AAW32313	C glutamicum prote
44	40	41.7	643	20	AAW16797	Amino acid sequenc
45	40	41.7	653	22	AAW02023	B. thuringiensis t
46	40	41.7	782	22	ABW21459	Novel human diagno
47	40	41.7	1151	20	AAW08711	Arabidopsis thalia
48	40	41.7	1186	22	AAW16796	Amino acid sequenc
49	40	41.7	1208	22	AAW02093	Bacillus thuringie
50	40	41.7	1209	22	AAW02094	Bacillus thuringie
51	40	41.7	1221	22	AAW00420	B. thuringiensis C
52	40	41.7	1221	22	AAW00421	B. thuringiensis C
53	40	41.7	1227	19	AAW44321	Bacillus thuringie
54	40	41.7	1227	20	AAW31990	Chimeric CryII ins
55	40	41.7	1227	22	AAW02046	B. thuringiensis t
56	40	41.7	1227	22	AAW19950	Bacillus thuringie
57	40	41.7	1228	15	AAW50955	Bacillus thuringie
58	40	41.7	1228	22	AAW84628	Amino acid sequenc
59	40	41.7	1228	22	AAW02039	B. thuringiensis t
60	40	41.7	1229	15	AAW54074	CryET5. Bacillus

ALIGNMENTS

RESULT	1
AAW42929	AAW42929 standard; peptide: 20 AA.
ID	AAW42929;
XX	AAW42929;
XX	28-Apr-1998 (first entry)
DT	
XX	
DE	Immunogenic Hepatitis A virus peptide YK-1331.
XX	
XX	Immunogenic peptide; immunogenic epitope: P2A protein;
KW	Immune response; antibody.
KW	
XX	
OS	Synthetic.
OS	Hepatitis A virus.
XX	
XX	W09740147-A1.
FN	
XX	
PD	30-OCt-1997.





Query Match 100.0%; Score 96; DB 7; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMIGLSGVORIKQ 20  
 ||||||||||||||||||  
 Db 961 kinladrmiglsyvgeikeq 980

## RESULT 6

AAR05697  
 ID AAR05697 standard; protein; 2227 AA.  
 XX AAR05697;

DT 15-AUG-1990 (first entry)  
 XX

DE Attenuated hepatitis A virus.

XX Hepatitis A virus; vaccine; attenuated.

OS Hepatitis A virus, strain HM-175.

XX Key Location/Qualifiers

FT 1..23

FT /label=VP4 = 1A

FT 24..245

FT /label=VP2 = 1B

FT 246..491

FT /label=VP3 = 1C

FT 492..791

FT /label=VP1 = 1D

FT 792..980

FT /label=2A

FT 981..1087

FT /label=2B

FT 1088..1422

FT /label=2C

FT 1423..1496

FT /label=3A

FT 1497..1519

FT /label=3B = VPg

FT 1520..1738

FT /label=3C

FT 1739..2227

FT /label=3D

XX Region

XX US4894228-A.

XX 16-JAN-1990.

XX 12-JUL-1988; 88US-0217824.

XX 12-JUL-1988; 88US-0217824.

XX 12-JUL-1988; 88US-0652967.

XX (USSH ) US DEPT HEALTH & HUMAN.

XX Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;

XX Daemer RJ, Gust ID;

XX WPI; 1990-075557/10.

XX N-PSDB; AAO03512.

XX Vaccine against hepatitis A virus infection - comprises novel

XX attenuated hepatitis A virus strain.

XX Claim 1; Fig 1; 18pp; English.

XX The attenuated HAV is useful for inducing protective immunity against

XX HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by

XX several nucleotide changes distributed throughout the genome, is

CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine.

SO Sequence 2227 AA;

Query Match 100.0%; Score 96; DB 11; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMIGLSGVORIKQ 20  
 ||||||||||||||||||  
 Db 961 kinladrmiglsyvgeikeq 980

## RESULT 7

AAW34074  
 ID AAW34074 standard; protein; 2227 AA.  
 XX AAW34074;

XX 27-APR-1998 (first entry)

DE Hepatitis A virus HM-175 protein sequence.

XX HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;

XX infection; vaccine.

XX Hepatitis A virus HM-175.

XX Key Location/Qualifiers

FT 1..23

FT /label= VP4

FT 24..245

FT /label= VP2

FT 246..491

FT /label= VP3

FT 492..791

FT /label= VP1

FT 792..980

FT /label= 2A

FT 981..1087

FT /label= 2B

FT 1088..1422

FT /label= 2C

FT 1423..1496

FT /label= 3A

FT 1497..1519

FT /label= 3B

FT 1520..1738

FT /label= 3C

FT 1739..2227

FT /label= 3D

XX Protein

XX WO97/40166-A2.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-US06506.

XX 19-APR-1996; 96US-0015642.

XX (USSH ) US SEC DEPT HEALTH.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Emerson SU, Purcell RH, Raychaudhuri G;

XX WPI; 1997-535850/49.

XX N-PSDB; AAT93023.

XX Human attenuated HAV genome containing simian HAV 2C gene - useful

XX as vaccines against HAV infection



XX Disclosure: Fig 13A-D; 66pp; English.  
PS  
XX  
CC This protein sequence is encoded by the human hepatitis A virus  
CC (HAV) HM-175 wild-type genome (see AF93023). Attenuated strain  
CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
CC kidney cells. A claimed DNA construct (1) comprises a genome of  
CC HAV, where the genome is a human attenuated HAV genome in which a  
CC region of the 2C gene has been replaced by a corresponding region  
CC from a 2C gene of a simian AGM-27 HAV genome (see AF93024). The  
CC region of the 2C gene from AGM-27 contained in the construct  
CC preferably encodes amino acids 120-328 of the 2C protein, amino  
CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
CC transcript of (1); (2) a cell transfected with (1) or the RNA  
CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
CC its RNA transcript, can be used as a vaccine for preventing HAV in  
CC a mammal. (1) or the RNA transcript can also be used to stimulate  
CC the production of protective antibodies in the mammal.  
XX  
SQ Sequence 2227 AA;  
  
Query Match 100.0%; Score 96; DB 18; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVOETKEQ 20  
DB 961 KlnLadrmLglsGvqelkeq 980  
|||||  
  
RESULT 8  
AAB18607  
ID AAB18607 standard; Protein; 2227 AA.  
XX  
AC AAB18607;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
XX  
KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
XX  
OS Hepatitis A virus.  
XX  
PN US6113912-A.  
XX  
PD 05-SEP-2000.  
XX  
PF 07-JUN-1995; 95US-0475886.  
XX  
PR 18-SEP-1992; 92US-0947338.  
PR 17-SEP-1993; 93WO-US08610.  
PR 10-MAR-1995; 95US-0397232.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
PI  
DR MPI: 2000-586464/55.  
DR N-PSDB; AAA75476.  
XX  
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type  
XX  
PS Disclosure: Fig 6A-K; 72pp; English.  
XX  
CC The present sequence is derived from a wild type hepatitis A virus  
CC (HAV) strain HM-174. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate

CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA;  
  
Query Match 100.0%; Score 96; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVOETKEQ 20  
DB 961 KlnLadrmLglsGvqelkeq 980  
|||||  
  
RESULT 9  
AAB18608  
ID AAB18608 standard; Protein; 2227 AA.  
XX  
AC AAB18608;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
XX  
KM HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
KM P-35 virus.  
XX  
OS Hepatitis A virus.  
XX  
PN US6113912-A.  
XX  
PD 05-SEP-2000.  
XX  
PF 07-JUN-1995; 95US-0475886.  
XX  
PR 18-SEP-1992; 92US-0947338.  
PR 17-SEP-1993; 93WO-US08610.  
PR 10-MAR-1995; 95US-0397232.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
PI  
DR MPI: 2000-586464/55.  
DR N-PSDB; AAA75477.  
XX  
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type  
XX  
PS Disclosure: Columns 67-78; 72pp; English.  
XX  
CC The present sequence is derived from passage 35 of a wild type  
CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX  
SQ Sequence 2227 AA;  
  
Query Match 100.0%; Score 96; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVOETKEQ 20  
DB 961 KlnLadrmLglsGvqelkeq 980  
|||||

XX		Proionibacterium acnes.
PN		M0200181581-A2.
PD		01-NOV-2001.
PX		20-APR-2001; 2001MO-USI2865.
PF		21-APR-2000; 2000US-1990A7P.
PR		02-JUN-2000; 2000US-2088A1P.
PP		07-JUL-2000; 2000US-2167A7P.
PA	(CORI-) CORIXA CORP.	
PI	Skeiky YAM, Persing DH, Mitcham JU., Wang SS,	Bhatia A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;	
DR	WPI; 2001-616774/71.	
DR	N-PSTB; AAS59515.	
PT	Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris -	
PS	Example 1; SEQ ID NO 3106; 1069pp; English.	
PX		
CC	<p>Sequences AAU39105-AAUG68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SARPO syndrome (synovitis), acne, pustulosis, hyperhidrosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).</p> <p>Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.</p>	
CC		
CC		
CC		
CC		
CQ	Sequence	80 AA:
	Query Match	44.8%; Score 43; DB 22; Length 80;
	Best Local Similarity	45.8%; Pred. No. 5.7;
	Matches 11; Conservative	5; Mismatches 4; Indels 4; Gaps 1;
OY	1 KINLADR---MTLSGVQEIKEQ 20	::::      :    ::   :
DG	51 eldvadrarraqlgpnvyddlaeq 74	:
	RESULT 12	
ID	AAG85023	
AC	AAG85023 standard; Protein; 922 AA.	
AA	AAG85023;	
DT	11-SEP-2001 (first entry)	
DE	Shrimp white spot Bacilliform virus (WSBV) protein 114.	
KM	Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection; antiviral agent; gene expression; antisense construct; transgenic viral resistant shrimp.	
KW		
XX		

OS	White spot syndrome virus.
XX	
PN	WO200138351-A2.
XX	
PD	31-MAY-2001.
XX	
PE	08-NOV-2000; 2000WO-US28886.
XX	
PR	24-NOV-1999; 99CN-0124717.
XX	
PA	(PENY-) PE CORP NY. (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANIC A. (SINO-) SINOGENOMAX CO LTD.
XX	
PI	Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX	
DR	WPI: 2001-355877/37.
XX	N-PDB: AAH62803.
PT	
PT	Primary nucleotide sequence of the shrimp white spot Bacilliform virus (WSBV), useful for producing viral polypeptides that can be used to screen for agents that are useful for treating WSBV infection -
XX	
PS	Claim 1; Figure 3; 626pp; English.
XX	
CC	The invention provides the primary nucleotide sequence of the WSBV genome (AAH62889), predicted transcript sequences (AAH62689-AAH62839) and encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences (AAH62840-63160) suitable for use as primers or probes. The nucleic acid molecules and proteins of the invention are useful for diagnosis and monitoring viral infection. In screens for antiviral agents and for controlling viral gene expression or activity during a treatment regimen. The nucleic acid molecules are also useful as antisense constructs to control viral gene expression in infected cells and tissues and to create transgenic viral resistant shrimp.
CC	
CC	
XX	Sequence 922 AA:
SQ	
Query Match	44.8%; Score 43; DB 22; Length 922; Best Local Similarity 50.0%; Pred. No. 98;
Matches	9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
OY	3 MLADRMGLSGVGQRKEQ 20    :       :    :
Db	388 nlrrktlglsdveekee 405
RESULT 13	
AAU49915	
ID	AAU49915 standard; Protein; 625 AA.
XX	
AC	AAU49915;
XX	
DT	27-FEB-2002 (first entry)
XX	
DE	Propionibacterium acnes immunogenic protein #10811.
XX	
KW	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis; uvulitis; endophthalmitis; bone joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.
XX	
OS	Propionibacterium acnes.
XX	
PN	WO200181581-A2.
XX	
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12865.
XX	
PR	21-APR-2000; 2000US-199047P. 02-JUN-2000; 2000US-208841P.
PR	

```

PR      07-JUL-2000; 2000US-216747P.
PA
PA      (CORI-) CORIXA CORP.
XX
PI      Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI      L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR      WPI; 2001-616774/71.
XX      N-PSDB; AAS59546.
XX
PT      Propionibacterium acnes polypeptides and nucleic acids useful for
PT      vaccinating against and diagnosing infections, especially useful for
PT      treating acne vulgaris -
XX
PS      Example 1; SEQ ID NO 11110; 1069pp; English.
XX
CC      Sequences AAU39105-AAU6017 represent Propionibacterium acnes immunogenic
CC      polypeptides. The proteins and their associated DNA sequences are used in
CC      the treatment, prevention and diagnosis of medical conditions caused by
CC      P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC      pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC      P. acnes is also involved in infections of bone, joints and the central
CC      nervous system, however it is particularly involved in the inflammatory
CC      lesions associated with acne vulgaris. A method for detecting the
CC      presence or absence of P. acnes in a patient comprises contacting a
CC      sample with a binding agent that binds to the proteins of the invention
CC      and determining the amount of bound protein in the sample. The
CC      polypeptides may be used as antigens in the production of antibodies
CC      specific for P. acnes proteins. These antibodies can be used to
CC      downregulate expression and activity of P. acnes polypeptides and
CC      therefore treat P. acnes infections. The antibodies may also be used as
CC      diagnostic agents for determining P. acnes presence, for example, by
CC      enzyme linked immunosorbent assay (ELISA).
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence      625 AA;

Query Match          43.8%; Score 42; DB 22; Length 625;
Best Local Similarity 64.3%; Pred. No. 93;
Matches    9; Conservative    2; Mismatches    3; Indels    0; Gaps    0;

QY      1 KINLADRMIGLSGV 14
       | |||||::||
Db      41 kstladrmigtcgv 54

RESULT 14
AAB60277
ID      AAB60277 standard; Protein: 1035 AA.
XX
XX      AAB60277;
DT      30-MAR-2001 (first entry)
XX
DE      Saccharomyces cerevisiae Cdc68 protein.
XX
KW      Cdc68 protein; Candida albicans; chromatin remodelling factor;
KW      DNA transcription; drug screening; inhibitor; antifungal agent;
KW      growth inhibition; degenerate PCR primer design.
OS      Saccharomyces cerevisiae.
XX
PN      WO200077215-A1.
XX
PD      21-DEC-2000.
XX
PF      15-JUN-2000; 2000WO-US16560.
XX
PR      16-JUN-1999; 99US-0139404.
XX

```



PF New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Disclosure; SEQ ID NO 4704; 21np + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (AB57737-AB572072).  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIRO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 213 AA;

Query Match 42.7%; Score 41; DB 22; Length 213;  
Best Local Similarity 46.7%; Pred. No. 40;  
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 4 LADRMGLSGVQDEIK 18  
||:|:|:|:|:|:  
DB 171 lasvlygagiqelr 185

RESULT 17  
ABG06041  
ID ABG06041 standard; Protein: 291 AA.  
XX  
AC ABG06041;  
XX  
DF 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6032.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS70228.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20: SEQ ID NO 36400; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIRO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 291 AA;

Query Match 42.7%; Score 41; DB 22; Length 291;  
Best Local Similarity 43.8%; Pred. No. 57;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 5 ADRMLGSGVQDEIKQ 20  
||:|:|:|:|:|:  
DB 263 atvtgagqkvevkg 278

RESULT 18  
AAY37809  
ID AAY37809 standard; Protein: 332 AA.  
XX  
AC AAY37809;  
XX  
DF 07-OCT-1999 (first entry)  
XX  
DE Protein which is specific to Chlamydia trachomatis.  
XX  
KM Vaccine; eye disease; conventional trachoma; nongonococcal trachoma;  
KM paratrachoma; inclusion conjunctivitis; genital disease; peritphalitis;  
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KM Bartholinitis; pneumonia; venereal lymphogranulomatosis.  
OS Chlamydia trachomatis.  
XX  
PN W09928475-A2.  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-IB01939.  
XX  
PR 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97ER-0015041.  
PR 17-DEC-1997; 97ER-0016034.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffiths R;  
PI  
DR WPI: 1999-371125/31.  
DR  
XX  
XX Genome sequence of Chlamydia trachomatis  
XX  
XX  
PS Disclosure: Page 1384-1385; 1755pp; English.  
XX  
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see A4201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nongonococcal trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC peritonitis, Bartholinitis; pneumonia in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.

XX Sequence 332 AA;

Query Match 42.7%; Score 41; DB 20; Length 332;  
Best Local Similarity 44.4%; Pred. No. 67;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 NLADRLGLSGVQEIKEQ 20  
Db 259 needtrialsgaglysee 276

## RESULT 19

AAW60855  
ID AAW60855 standard; Protein; 366 AA.

AC AAW60855;

DT 01-OCT-1998 (first entry)

DE Mouse CD14 protein.

CD14; B cell activator; bovine lactation-associated immunotropic protein;  
LAIR; T cell immunodeficiency; X-linked hyper-IgM syndrome; allergy;  
common variable immunodeficiency; X-linked agammaglobulinemia; vaccine;  
infant feeding formulae; mouse.

Mus sp.

WO9822580-A2.

28-MAY-1998.

18-NOV-1997; 97WO-CA00880.

18-NOV-1996; 96US-0746883.

(WELL-) WELLESLEY HOSPITAL FOUND.

Alizadeh-Khiavi K, Filipp D, Julius MH;

WPI; 1998-312466/27.

N-PSDB; AAV37229.

New bovine polypeptide that activates mammalian B cell(s) - used  
e.g. to treat T cell immunodeficiency or allergy, as vaccine  
adjuvant, as T cell surrogate for infants, and for monoclonal  
antibody production, also specific antibodies for treating B cell  
hyperactivity

Claim 14; Fig 7; 64pp; English.

This sequence is the mouse CD14 protein of the invention. The CD14  
protein was used to isolate the bovine CD14 of the invention,  
which is able to activate mammalian B cells. The protein is also known  
as bovine lactation-associated immunotropic protein (LAIR), and is used  
to activate B cells, particularly in humans. Particularly it is  
administered to subjects: (a) with CD40 negative or deficient B cells;  
(b) suffering from T cell immunodeficiency (e.g. X-linked hyper-IgM  
syndrome, common variable immunodeficiency or X-linked  
agammaglobulinemia) or allergy (i.e. with CD40 ligand negative or  
defective T cells); or (c) to induce growth and differentiation of  
B cells to highly productive Ig secreting cells. Particular applications  
are in infant feeding formulae (as immunostimulant) and as adjuvant in  
vaccines (optionally with bovine CD14 coupled to the antigen). The DNA  
sequences are also used to enrich mammalian B cells secreting a  
monoclonal antibody (MAb) of particular antigenic specificity, by  
activating cells with sub-optimal amount of the DNA in combination with

CC the antigen. The enriched B cells are then used to produce hybridomas  
CC that produce specific MAb. Antibodies raised against human CD14 are used  
CC to reduce/inhibit activity of B cells that are hyperactivated by high  
CC serum levels of CD14. Bovine CD14 stimulates growth (induce DNA  
CC synthesis) in resting murine spleen cells and is 200 times more  
CC effective than lipopolysaccharide (LPS), with the effect unaffected by  
CC presence of serum. It also induces Ig secretion and a partial isotype  
CC switch from IgM to IgG, in absence of T cells.

XX Sequence 366 AA;

Query Match 42.7%; Score 41; DB 19; Length 366;  
Best Local Similarity 63.6%; Pred. No. 75;  
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMGLSGVQEI 17  
Db 107 rvi1sg1gel 117

## RESULT 20

AAV53879  
ID AAV53879 standard; Protein; 366 AA.

AC AAV53879;

DT 13-MAR-2000 (first entry)

DE A murine CD14 protein which is used to treat sepsis.

CD14; sepsis; epithelial cell; defensin; antibiotic polypeptide;  
B cell; infection; immune deficient; wound healing; burn; ulcer;  
transgenic animal.

Mus sp.

WO9961468-A2.

02-DEC-1999.

27-MAY-1999; 99WO-CA00482.

27-MAY-1998; 98US-0086884.

(GEMM-) GEMMA BIOTECHNOLOGY LTD.

Julius MH, Filipp D;

WPI; 2000-072604/06.

N-PSDB; AAZ6820.

Treating sepsis and activating B cells by administration of soluble  
CD14, e.g. for preventing infection

Claim 10; Page 59-60; 61pp; English.

The present sequence represents a CD14 protein. The protein is used in  
the method of the invention. The specification describes a method  
whereby symptoms of sepsis are ameliorated by exposing epithelial cells  
in a mammal directly to soluble CD14 or its fragments. The CD14 is  
capable of stimulating expression of a defensin by the epithelial cells.  
CD14 directly induces expression of antibiotic polypeptides,  
specifically defensins, in mammals and stimulates B cells (i.e. induces  
differentiation of their precursors). CD14 protein is used to treat  
sepsis; to stimulate B cells, particularly for protection against  
infection (viral, bacterial, fungal or yeast), especially in immune  
deficient subjects, and to promote healing of wounds (burns, ulcers).  
Transgenic animals containing exogenous sequences encoding CD14 are  
used for enhanced production of CD14 in the milk.

XX Sequence 366 AA;

Query Match 42.7%; Score 41; DB 21; Length 366;  
Best Local Similarity 63.6%; Pred. No. 75;  
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMLGSGVOEI 17  
1:1:1:1:1:1:  
Db 107 rvlglsglql 117

## RESULT 21

AAU36494  
ID AAU36494 standard; Protein: 605 AA.

AC AAU36494;

DT 14-FEB-2002 (first entry)

DE Pseudomonas aeruginosa cellular proliferation protein #484.

KW Antisense: prokaryotic cellular proliferation protein;

KM antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001MO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlson KL, Zyakind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

DR N-PSDB: AAS54353.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 12087; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 605 AA;

SO

Query Match 42.7%; Score 41; DB 22; Length 605;  
Best Local Similarity 47.4%; Pred. No. 13e+02;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIK 19  
1:1:1:1:1:1:  
Db 18 ktllvdrlklsgldrke 36

## RESULT 22

AAW69844  
ID AAW69844 standard; Protein: 708 AA.

AC AAW69844;

DT 07-OCT-1998 (first entry)

DE Amino acid sequence of p71, a protein present in SA-17S complex.

XX p71 protein; secretion associated 17S complex; SA-17S;

KW syntaxin-containing complex; SC complex; screening; modulate;

KW vesicular release; synaptic transmission; secretory process;

KW treatment; affective disorder; depression; manic-depressive disorder;

KW anxiety disorder; neurodegenerative disease; schizophrenia; anaesthesia;

KW hormonal imbalance; antigen processing; ss.

OS Rattus sp.

XX WO9828419-A2.

XX 02-JUL-1998.

PF 19-DEC-1997; 97MO-US23498.

PR 20-DEC-1996; 96US-0033905.

XX (STRD ) UNIV LELAND STANFORD JUNIOR.

PI Scheller RH;

XX WPI: 1998-377650/32.

DR N-PSDB: AAW42649.

XX New isolated vesicle secretion associated poly(peptide(s) - used to

PT develop products for treating e.g. effective disorder,

PT neurodegenerative disease, hormone imbalances, immune system

PT disorders or tumours

PS Claim 1a; Pages 85-87; 133pp; English.

XX The present sequence represents a p71 protein. This protein is present

CC the secretion associated 17S (SA-17S) complex. Eight proteins form the

CC SA-17S complex, which binds a syntaxin-containing (SC) complex. The

CC SA-17S polypeptides and nucleotide sequences encoding them can be used

CC for screening for compounds which modulate vesicular release involved in

CC synaptic transmission and other secretory processes. Compounds which

CC enhance binding between the SA-17S and SC complexes may be used to treat

CC an affective disorder such as depression, manic-depressive disorders and

CC anxiety disorders, or a neurodegenerative disease such as Parkinson's

CC disease or Huntington's disease. Compounds which inhibit binding between

CC the SA-17S and SC complexes may be used to treat a disorder of thought,

CC such as schizophrenia, or for anaesthesia. The compounds can also be used

CC to intervene in the endocrine system for treatment of hormonal

CC imbalances, the immune system for intervention in antigen processing,

CC secreted immunomodulators, and viral processing, as well as

CC anti-tumour applications, such as regulation of membrane trafficking

CC during rapid cell division.

XX Sequence 708 AA;

SO





CC AAW50145), or cyanobacteria such as Synechocystis. Truncated  
 CC apoproteins consisting of the N-terminal chromophore domain are  
 CC especially preferred. Recombinant apoproteins assemble  
 CC spontaneously with the bilin chromophore. Claimed compositions  
 CC comprise a protein, glycoprotein, antibody or nucleic acid to be  
 CC detected linked to the fluorescent adduct. They are used in assays  
 CC for detecting the other member of a specific binding pair, e.g.  
 CC immunoassay of antigens, immuno-histochemical labelling, as nucleic  
 CC acid probes for Southern blotting, for identification of manufactured  
 CC products, also to detect protein-protein interactions, including  
 CC studies on intracellular protein localisation and identification of  
 CC transfected cells. The phytofluors make ideal fluorescent markers  
 CC because they have a long wavelength absorption maximum and high  
 CC molar absorption coefficient, and are stable to light and pH.

CC Sequence 748 AA;  
 XX

Query Match 42.7%; Score 41; DB 19; Length 748;  
 Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 NIADRMGLSGVOE 16  
 | ||||| : ||  
 Db 354 nhpor11gltsgsq 367

# RESULT 25

ABG19753  
 ID ABG19753 standard; Protein; 92 AA.

AC ABG19753;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #19744.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS83940.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID No 50112; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SO Sequence 92 AA;

Query Match 41.7%; Score 40; DB 22; Length 92;  
 Best Local Similarity 61.5%; Pred. No. 23;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 7 RMGLSGVOEIKE 19  
 | : ||||| : ||  
 Db 62 r11gltsgaehve 74

Search completed: June 16, 2002, 00:01:49  
 Job time: 12738 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:03:20 ; Search time 79.04 Seconds  
(without alignments)  
6.181 Million cell updates/sec

Title: US-09-171-432a-46

Perfect score: 96

Sequence: 1 KINLADRMGLSGVOEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2227	3 US-08-475-886-2	Sequence 2, Appli
2	96	100.0	2227	3 US-08-475-886-4	Sequence 4, Appli
3	96	100.0	2227	3 US-08-475-886-6	Sequence 6, Appli
4	96	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
5	96	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
6	96	100.0	2227	4 US-09-171-387-2	Sequence 2, Appli
7	42	43.8	1250	1 US-08-441-139-9	Sequence 9, Appli
8	41	42.7	365	3 US-08-746-883-6	Sequence 6, Appli
9	40	41.7	332	1 US-08-469-649-2	Sequence 2, Appli
10	40	41.7	638	1 US-08-846-762-95	Sequence 95, Appli
11	40	41.7	643	4 US-09-178-252-25	Sequence 23, Appli
12	40	41.7	1186	4 US-09-178-252-23	Sequence 7, Appli
13	40	41.7	1207	1 US-07-951-715A-7	Sequence 7, Appli
14	40	41.7	1207	2 US-08-459-448A-7	Sequence 7, Appli
15	40	41.7	1207	3 US-08-459-595A-7	Sequence 7, Appli
16	40	41.7	1207	3 US-08-459-504B-7	Sequence 7, Appli
17	40	41.7	1207	3 US-08-459-444-7	Sequence 7, Appli
18	40	41.7	1207	3 US-09-053-549-8	Sequence 8, Appli
19	40	41.7	1207	4 US-09-547-422-7	Sequence 8, Appli
20	40	41.7	1227	1 US-08-448-170-8	Sequence 8, Appli
21	40	41.7	1227	3 US-09-053-549-2	Sequence 9, Appli
22	40	41.7	1227	4 US-08-961-803-9	Sequence 9, Appli
23	40	41.7	1229	1 US-08-100-709-4	Sequence 4, Appli
24	40	41.7	1229	1 US-08-176-865-4	Sequence 4, Appli
25	40	41.7	1229	1 US-08-474-038-4	Sequence 4, Appli
26	40	41.7	1229	2 US-08-779-046-4	Sequence 4, Appli
27	40	41.7	1229	2 US-08-881-340-4	Sequence 4, Appli

28	40	41.7	1388	2 US-08-685-576-1	Sequence 1, Appli
29	39	40.6	528	3 US-08-904-871-5	Sequence 5, Appli
30	39	40.6	748	3 US-08-904-871-6	Sequence 6, Appli
31	39	40.6	748	3 US-08-904-871-13	Sequence 13, Appli
32	38	39.6	185	2 US-08-691-814B-12	Sequence 12, Appli
33	38	39.6	261	1 US-07-940-605A-2	Sequence 2, Appli
34	38	39.6	261	1 US-08-184-422-8	Sequence 8, Appli
35	38	39.6	261	1 US-08-360-923A-2	Sequence 2, Appli
36	38	39.6	261	1 US-08-446-922-4	Sequence 4, Appli
37	38	39.6	261	1 US-08-431-055-4	Sequence 4, Appli
38	38	39.6	261	2 US-08-690-036-2	Sequence 2, Appli
39	38	39.6	261	2 US-08-249-189-12	Sequence 12, Appli
40	38	39.6	261	2 US-08-484-624A-12	Sequence 12, Appli
41	38	39.6	261	2 US-08-477-733B-12	Sequence 12, Appli
42	38	39.6	261	2 US-08-763-995-2	Sequence 2, Appli
43	38	39.6	261	3 US-09-088-913A-12	Sequence 12, Appli
44	38	39.6	261	3 US-08-588-771B-8	Sequence 8, Appli
45	38	39.6	261	4 US-08-769-819-12	Sequence 12, Appli
46	38	39.6	261	4 US-08-770-974-12	Sequence 12, Appli
47	38	39.6	261	4 US-08-858-197-4	Sequence 4, Appli
48	38	39.6	261	5 PCT-US93-10034-4	Sequence 4, Appli
49	38	39.6	273	1 US-08-446-922-11	Sequence 11, Appli
50	38	39.6	273	2 US-08-249-189-21	Sequence 21, Appli
51	38	39.6	273	2 US-08-484-624A-21	Sequence 21, Appli
52	38	39.6	273	2 US-08-477-733B-21	Sequence 21, Appli
53	38	39.6	273	3 US-09-088-913A-21	Sequence 21, Appli
54	38	39.6	273	4 US-08-769-819-21	Sequence 21, Appli
55	38	39.6	273	4 US-08-770-974-21	Sequence 21, Appli
56	38	39.6	420	3 US-08-685-871-58	Sequence 38, Appli
57	38	39.6	456	2 US-08-910-731-2	Sequence 2, Appli
58	38	39.6	456	2 US-08-795-395-2	Sequence 2, Appli
59	38	39.6	473	2 US-08-249-189-16	Sequence 16, Appli
60	38	39.6	473	2 US-08-484-624A-16	Sequence 16, Appli

## ALIGNMENTS

RESULT 1  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PORCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475, 886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 96; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KINLADRMGLSGVOEIKEQ 20  
|||||  
Db 961 KINLADRMGLSGVOEIKEQ 980

```
RESULT 2
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match          100.0%; Score 96; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20
|||||
DB 961 KINLADRMGLSGVQEIKEQ 980

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match          100.0%; Score 96; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20
|||||
DB 961 KINLADRMGLSGVQEIKEQ 980

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHADHURI, GOPA;
```

```
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

Query Match          100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20
|||||
DB 961 KINLADRMGLSGVQEIKEQ 980

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match          100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20
|||||
DB 961 KINLADRMGLSGVQEIKEQ 980

RESULT 6
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHADHURI, GOPA;
```

EMERSON, SUZANNE, U. ;  
PURCELL, ROBERT, H. ;  
TITLE OF INVENTION: SIMIAN-HUMAN HAV ;  
HAVING A CHIMERIC 2C PROTEIN ;  
NUMBER OF SEQUENCES: 3 ;  
CORRESPONDENCE ADDRESS: ;  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P. ;  
STREET: 345 PARK AVENUE ;  
CITY: NEW YORK ;  
STATE: NEW YORK ;  
COUNTRY: USA ;  
ZIP: 10154 ;  
COMPUTER READABLE FORM: ;  
MEDIUM TYPE: FLOPPY DISK ;  
COMPUTER: IBM PC COMPATIBLE ;  
OPERATING SYSTEM: PC-DOS/MS-DOS ;  
SOFTWARE: MICROSOFT WORD 97 ;  
CURRENT APPLICATION DATA: ;  
APPLICATION NUMBER: US/09/171,387 ;  
FILING DATE: 24-Mar-1999 ;  
PRIOR APPLICATION NUMBER: PCT/US97/06506 ;  
FILING DATE: 18-APR-1997 ;  
APPLICATION NUMBER: US60/015,642 ;  
FILING DATE: 19-APR-1996 ;  
ATTORNEY/AGENT INFORMATION: ;  
NAME: William S. Feller ;  
REGISTRATION NUMBER: 26,728 ;  
REFERENCE/DOCKET NUMBER: 2026-4229051 ;  
TELECOMMUNICATION INFORMATION: ;  
TELEPHONE: (212) 758-4800 ;  
TELEFAX: (212) 751-6849 ;  
INFORMATION FOR SEQ ID NO: 2 ;  
SEQUENCE CHARACTERISTICS: ;  
LENGTH: 2227 amino acids ;  
TYPE: amino acid ;  
STRANDEDNESS: unknown ;  
TOPOLOGY: unknown ;  
SEQUENCE DESCRIPTION: SEQ ID NO: 2 ;  
US-09-171-387-2 ;  
Query Match 100.0%; Score 96; DB 4; Length 2227 ;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08 ;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0 ;  
QY 1 KINLADRMGLSGVOEIKRQ 20 ;  
Db 961 KINLADRMGLSGVOEIKRQ 980 ;  
RESULT 7 ;  
US-08-441-139-9 ;  
Sequence 9, Application US/08441139 ;  
Patent No. 5773245 ;  
GENERAL INFORMATION: ;  
APPLICANT: Wittnup, Dr. Karl D. ;  
APPLICANT: Robinson, Anne S. ;  
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF ;  
TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS ;  
NUMBER OF SEQUENCES: 20 ;  
CORRESPONDENCE ADDRESS: ;  
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER ;  
STREET: 400 Garden City Plaza ;  
CITY: Garden City ;  
STATE: NY ;  
COUNTRY: USA ;  
ZIP: 11530 ;  
COMPUTER READABLE FORM: ;  
MEDIUM TYPE: FLOPPY disk ;  
COMPUTER: IBM PC compatible ;  
OPERATING SYSTEM: PC-DOS/MS-DOS ;  
SOFTWARE: PatentIn Release #1.0, Version #1.25 ;

CURRENT APPLICATION DATA: ;  
APPLICATION NUMBER: US/08/441,139 ;  
FILING DATE: 15-MAY-1995 ;  
CLASSIFICATION: 435 ;  
PRIOR APPLICATION DATA: ;  
APPLICATION NUMBER: US 08/089,997 ;  
FILING DATE: 06-JUL-1993 ;  
ATTORNEY/AGENT INFORMATION: ;  
NAME: Digiglio, Frank S. ;  
REGISTRATION NUMBER: 31,346 ;  
REFERENCE/DOCKET NUMBER: 8646 ;  
TELECOMMUNICATION INFORMATION: ;  
TELEPHONE: 516-742-4343 ;  
TELEFAX: 516-742-4366 ;  
TELEX: 230 901 SANS UR ;  
INFORMATION FOR SEQ ID NO: 9 ;  
SEQUENCE CHARACTERISTICS: ;  
LENGTH: 1250 amino acids ;  
TYPE: amino acid ;  
TOPOLOGY: linear ;  
MOLECULE TYPE: protein ;  
US-08-441-139-9 ;  
Query Match 43.8%; Score 42; DB 1; Length 1250 ;  
Best Local Similarity 42.1%; Pred. No. 39 ;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0 ;  
QY 2 INLADRMGLSGVOEIKRQ 20 ;  
Db 871 INLADRMGLSGVOEIKRQ 889 ;  
RESULT 8 ;  
US-08-746-883-6 ;  
Sequence 6, Application US/08746883 ;  
Patent No. 6093693 ;  
GENERAL INFORMATION: ;  
APPLICANT: Julius, Michael H., Philipp, Dominic, ;  
APPLICANT: Alizadeh-Khivari, Kamel ;  
TITLE OF INVENTION: B Cell Activation ;  
NUMBER OF SEQUENCES: 13 ;  
CORRESPONDENCE ADDRESS: ;  
ADDRESSEE: Blake, Cassels & Graydon ;  
STREET: Box 25, Commerce Court West ;  
CITY: Toronto ;  
COUNTRY: Canada ;  
ZIP: M5L 1A9 ;  
COMPUTER READABLE FORM: ;  
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage ;  
COMPUTER: COMPAQ, IBM PC compatible ;  
OPERATING SYSTEM: MS-DOS 5.1 ;  
SOFTWARE: WORD PERFECT ;  
CURRENT APPLICATION DATA: ;  
APPLICATION NUMBER: US/08/746,883 ;  
FILING DATE: No. 6093693ember 18, 1996 ;  
ATTORNEY/AGENT INFORMATION: ;  
NAME: Hunt, John C. ;  
REGISTRATION NUMBER: 36,424 ;  
REFERENCE/DOCKET NUMBER: 47841/00008 ;  
TELECOMMUNICATION INFORMATION: ;  
TELEPHONE: (416) 863-4344 ;  
TELEFAX: (416) 863-2653 ;  
INFORMATION FOR SEQ ID NO: 6 ;  
SEQUENCE CHARACTERISTICS: ;  
LENGTH: 366 amino acids ;  
TYPE: amino acid ;  
STRANDEDNESS: single ;  
TOPOLOGY: linear ;  
US-08-746-883-6 ;  
Query Match 42.7%; Score 41; DB 3; Length 366 ;

Best Local Similarity 63.6%; Pred. No. 13;  
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMJGSGVOEI 17  
1:1:1:1:1:1:  
Db 107 RVJGSGLOEL 117

RESULT 9

US-08-469-649-2  
; Sequence 2, Application US/08469649  
; Patent No. 5728562

; GENERAL INFORMATION:

; APPLICANT: Shigyo, Tatsuro

; APPLICANT: Sugihara, Kohji

; APPLICANT: Takamoto, Yoji

; APPLICANT: Takashio, Masachika

; APPLICANT: Kamimura, Minoru

; APPLICANT: Yamamoto, Kazumi

; APPLICANT: Kojima, Yoshio

; APPLICANT: Kikuchi, Toshiro

; APPLICANT: Eml, Shigenori

; TITLE OF INVENTION: AN ISOLATED URICASE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Frischauf, Holtz, Goodman, Langer &

; ADDRESSEE: Chick, P.C.

; STREET: 767 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10017-2023

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5, 1/4 inch, 360 Kb Storage

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,649

; FILING DATE: June 6, 1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Barth, Richard

; REGISTRATION NUMBER: 28,180

; REFERENCE/DOCKET NUMBER: 890578CIPD

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 319-4900

; TELEFAX: (212) 319-5101

; TELEX: 236268

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 332 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-469-649-2

Query Match 41.7%; Score 40; DB 1; Length 332;  
Best Local Similarity 41.2%; Pred. No. 18;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 INIADRMJGSGVOEIK 18  
1:1:1:1:1:1:1:  
Db 162 LNTFQOQSGIAGIQLEIK 178

RESULT 10

US-08-846-762-95

; Sequence 95, Application US/08846762A  
; Patent No. 5994072

; GENERAL INFORMATION:

; APPLICANT: Lam, Joseph S.

; APPLICANT: Burrows, Lori  
; APPLICANT: Charter, Deborah  
; APPLICANT: de Kievit, Teresa  
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly  
; TITLE OF INVENTION: of O-antigen in Pseudomonas aeruginosa

; FILE REFERENCE: 6580-089

; CURRENT APPLICATION NUMBER: US/08/846,762A

; CURRENT FILING DATE: 1997-04-30

; NUMBER OF SEQ ID NOS: 100

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 95

; LENGTH: 638

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-08-846-762-95

Query Match 41.7%; Score 40; DB 2; Length 638;  
Best Local Similarity 47.4%; Pred. No. 40;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 INIADRMJGSGVOEIKQ 20  
1:1:1:1:1:1:1:  
Db 515 IDIAKRMJGSGIKSEE 533

RESULT 11

US-09-178-252-25

; Sequence 25, Application US/09178252  
; Patent No. 6218188

; GENERAL INFORMATION:

; APPLICANT: Cardineau, Guy A.

; APPLICANT: Stelman, Steven J.

; APPLICANT: Narva, Kenneth E.

; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins

; FILE REFERENCE: MA-714XC2

; CURRENT APPLICATION NUMBER: US/09/178,252

; CURRENT FILING DATE: 1998-10-23

; EARLIER APPLICATION NUMBER: 60/065,215

; EARLIER FILING DATE: 1997-11-12

; EARLIER APPLICATION NUMBER: 60/076,445

; EARLIER FILING DATE: 1998-03-02

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 25

; LENGTH: 643

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Toxin encoded by synthetic B.t. gene

; OTHER INFORMATION: Toxin encoded by synthetic B.t. gene

US-09-178-252-25

Query Match 41.7%; Score 40; DB 4; Length 643;  
Best Local Similarity 61.5%; Pred. No. 40;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INIADRMJGSGV 14  
1:1:1:1:1:1:1:  
Db 60 INIAGRILGVLGV 72

RESULT 12

US-09-178-252-23

; Sequence 23, Application US/09178252  
; Patent No. 6218188

; GENERAL INFORMATION:

; APPLICANT: Cardineau, Guy A.

; APPLICANT: Stelman, Steven J.

; APPLICANT: Narva, Kenneth E.

; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins

FILE REFERENCE: MA-714XC2  
CURRENT APPLICATION NUMBER: US/09/178,252  
CURRENT FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 60/065,215  
EARLIER FILING DATE: 1997-11-12  
EARLIER APPLICATION NUMBER: 60/076,445  
EARLIER FILING DATE: 1998-03-02  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 23  
LENGTH: 1186  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Toxin encoded by synthetic B.t. gene  
US-09-178-252-23

Query Match 41.7%; Score 40; DB 4; Length 1186;  
Best Local Similarity 61.5%; Pred. No. 84;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMIGLSGV 14  
||| | | | | |  
Db 60 INIAGRILGVLCV 72

RESULT 13  
US-07-951-715A-7  
Sequence 7, Application US/07951715A  
Patent No. 5625136  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalin M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttle, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/951,715A  
FILING DATE: 25-SEP-1992  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Spill, W. Murray  
REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8615  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-951-715A-7

Query Match 41.7%; Score 40; DB 1; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMIGLSGV 14  
||| | | | | |  
Db 34 INIAGRILGVLCV 46

RESULT 14  
US-08-459-448A-7  
Sequence 7, Application US/08459448A  
Patent No. 5859336  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalin M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttle, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5859336artis Corporation  
STREET: Patent & Trademark Dept., 520 White Plains  
STREET: Rd., POB 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,448A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-448A-7

Query Match 41.7%; Score 40; DB 2; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMIGLSGV 14  
|||1:|:|1|  
Db 34 INIAGRILGVLVG 46

RESULT 15  
US-08-459-595A-7  
Sequence 7, Application US/08459595A  
Patent No. 6018104  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalin M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6018104artis Corporation  
STREET: Patent & Trademark Dept., 520 White Plains  
STREET: Rd., POB 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-10S  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,595A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-595A-7

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMIGLSGV 14  
|||1:|:|1|  
Db 34 INIAGRILGVLVG 46

RESULT 16  
US-08-459-504B-7  
Sequence 7, Application US/08459504B  
Patent No. 6075185  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalin M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6075185artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,504B  
FILING DATE: 02-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/459,595  
FILING DATE: 02-JUN-1995  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Melgs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV  
TELECOMMUNICATION INFORMATION:



TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-504B-7

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. NO. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMGLSGV 14  
||:|:|:|:|  
Db 34 INIAGRIIGVLGV 46

## RESULT 17

US-08-459-444-7  
Sequence 7, Application US/08459444A  
Patent No. 6121014

## GENERAL INFORMATION:

APPLICANT: Koziel, Michael G.  
Desai, Nalini M.  
Lewis, Kelly S.  
Kramer, Vance C.  
Warren, Gregory W.  
Evola, Stephen V.  
Crossland, Lyle D.  
Wright, Martha S.  
Merlin, Ellis J.  
Launis, Karen L.

TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED  
NUCLEIC ACID CODING SEQUENCE

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.  
STREET: 3054 Cornwallis Road

CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,444A  
FILING DATE: 02-Jun-1995

## CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991

## ATTORNEY/AGENT INFORMATION:

NAME: Melis, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689

## INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-08-459-444-7

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. NO. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMGLSGV 14  
||:|:|:|:|  
Db 34 INIAGRIIGVLGV 46

## RESULT 18

US-09-053-549-8  
Sequence 8, Application US/09053549  
Patent No. 6121521

## GENERAL INFORMATION:

APPLICANT: Desai, Nalini  
TITLE OF INVENTION: NO. 6121521el Insecticidal Protein and Gene  
NUMBER OF SEQUENCES: 8

## CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6121521artis Corporation  
STREET: 3054 Cornwallis Rd.  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,549  
FILING DATE: 01-APR-1998

## CLASSIFICATION: 800

## ATTORNEY/AGENT INFORMATION:

NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1995  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689

## INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-053-549-8

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. NO. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 INLADRMGLSGV 14  
||:|:|:|:|  
Db 34 INIAGRIIGVLGV 46

## RESULT 19

US-09-547-422-7  
Sequence 7, Application US/09547422  
Patent No. 6320100

## GENERAL INFORMATION:

APPLICANT: Koziel, Michael G.  
Desai, Nalini M.  
Lewis, Kelly S.  
Kramer, Vance C.  
Warren, Gregory W.  
Evola, Stephen V.  
Crossland, Lyle D.

Wright, Martha S.  
Merlin, Ellis J.  
Launis, Karen L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6320100arts Agribusiness Biotechnology Research, Inc.  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/547,422  
FILING DATE: 11-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/459,595  
FILING DATE: 02-JUN-1995  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,341  
REFERENCE/DOCKET NUMBER: S-18605H  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-547-422-7

Query Match 41.7%; Score 40; DB 4; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|  
Db 34 INIAGRILGLV 46

RESULT 20  
US-08-448-170-8  
Sequence 8, Application US/08448170  
Patent No. 5723758  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Cummings, David A.  
APPLICANT: Cannon, Raymond J.C.  
APPLICANT: Naraya, Kenneth E.  
APPLICANT: Steiman, Steve  
TITLE OF INVENTION: No. 5723758el Bacillus thuringiensis Isolate Denoted  
TITLE OF INVENTION: B.c. p515862, Active Against Lepidopteran Pests, and Genes  
TITLE OF INVENTION: Encoding Lepidopteran-Active Toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville

STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,170  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/069,902  
FILING DATE: 01-JUNE-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/759,247  
FILING DATE: 13-SEPT-1991  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/S 102D.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-448-170-8

Query Match 41.7%; Score 40; DB 1; Length 1227;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|  
Db 60 INIAGRILGLV 72

RESULT 21  
US-09-053-549-2  
Sequence 2, Application US/09053549  
Patent No. 6121521  
GENERAL INFORMATION:  
APPLICANT: Desai, Nalini  
TITLE OF INVENTION: No. 6121521el Insecticidal Protein and Gene  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6121521arts Corporation  
STREET: 3054 Cornwallis Rd.  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,549  
FILING DATE: 01-APR-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403

REFERENCE/DOCKET NUMBER: CGC 1995  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-053-549-2

Query Match 41.7%; Score 40; DB 3; Length 1227;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
DB 55 INIAGRILGLV 67

RESULT 22  
US-08-961-803-9  
Sequence 9, Application US/08961803  
Patent No. 6150589  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Cummings, David A.  
APPLICANT: Cannon, Raymond J.C.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Steiman, Steve  
TITLE OF INVENTION: NO. 6150589el Bacillus thuringiensis Isolate Denoted  
TITLE OF INVENTION: B. t. PSI58C2, Active Against Lepidopteran Pests, and Genes  
TITLE OF INVENTION: Encoding Lepidopteran-Active Toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jay M. Sanders  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,803  
FILING DATE: 31-OCT-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/069,902  
FILING DATE: 01-JUNE-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/759,247  
FILING DATE: 13-SEPT-1991  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/448,170  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M. 39,355  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: M/S 102DCD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1227 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-961-803-9

Query Match 41.7%; Score 40; DB 4; Length 1227;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
DB 60 INIAGRILGLV 72

RESULT 23  
US-08-100-709-4  
Sequence 4, Application US/08100709  
Patent No. 5322687  
GENERAL INFORMATION:  
APPLICANT: Donovan, William P.  
APPLICANT: Tan, Yuding  
APPLICANT: Jany, Christine S.  
APPLICANT: Gonzalez Jr., Jose M.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5  
TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Panlitch Schwarze Jacobs & Nadel c/o A.S.  
ADDRESS: Nadel  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/100,709  
FILING DATE: 19930729  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Egoft, Christopher  
REGISTRATION NUMBER: 27633  
REFERENCE/DOCKET NUMBER: 7205-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-757-1590  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1229 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-100-709-4

Query Match 41.7%; Score 40; DB 1; Length 1229;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
DB 60 INIAGRILGLV 72

RESULT 24  
US-08-176-865-4  
Sequence 4, Application US/08176865

Patent No. 5616319  
GENERAL INFORMATION:  
APPLICANT: Donovan, William P.  
APPLICANT: Tan, Yuhling  
APPLICANT: Jany, Christine S.  
APPLICANT: Gonzalez Jr., Jose M.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS CYET4 AND CYET5  
TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.  
ADDRESSEE: Nadel  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/176,865  
FILING DATE: 30-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/100,709  
FILING DATE: 29-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Egolf, Christopher  
REGISTRATION NUMBER: 27633  
REFERENCE/DOCKET NUMBER: 7205-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-757-1590  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1229 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-176-865-4

Query Match 41.7%; Score 40; DB 1; Length 1229;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|:|  
Db 60 INIAGRILGLGV 72

RESULT 25  
US-08-474-038-4  
Sequence 4, Application US/08474038  
Patent No. 5679343  
GENERAL INFORMATION:  
APPLICANT: Donovan, William P.  
APPLICANT: Tan, Yuhling  
APPLICANT: Jany, Christine S.  
APPLICANT: Gonzalez Jr., Jose M.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS CYET4 AND CYET5  
TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.  
ADDRESSEE: Nadel  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,038  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/176,865  
FILING DATE: 30-DEC-1993  
APPLICATION NUMBER: US 08/100,709  
FILING DATE: 29-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Egolf, Christopher  
REGISTRATION NUMBER: 27633  
REFERENCE/DOCKET NUMBER: 7205-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-757-1590  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1229 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-038-4

Query Match 41.7%; Score 40; DB 1; Length 1229;  
Best Local Similarity 61.5%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|:|  
Db 60 INIAGRILGLGV 72

Search completed: June 16, 2002, 00:03:21  
Job time: 8535 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:05:23 ; Search time 108.75 Seconds

(without alignments)  
17.672 Million cell updates/sec

Title: US-09-171-432a-46

Perfect score: 96

Sequence: 1 KINLADRMGLSGVQEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2227	1 GNNYHM	genome polyprotein
2	96	100.0	2227	1 GNNYHM	genome polyprotein
3	96	100.0	2227	1 GNNYHM	genome polyprotein
4	96	100.0	2227	1 GNNYHM	genome polyprotein
5	96	100.0	2227	1 GNNYHM	genome polyprotein
6	47	49.0	145	2 B25199	heat shock 16k pro
7	47	49.0	172	2 S33416	heat shock protein
8	44	45.8	414	2 T06303	enoyl-CoA hydratase
9	44	45.8	646	2 S72609	GTP-binding membra
10	44	45.8	653	2 G70683	probable lepa - My
11	44	45.8	788	2 S67595	hypothetical prote
12	43	44.8	145	2 B24289	heat shock protein
13	43	44.8	180	2 S64321	hypothetical prote
14	43	44.8	180	2 AC0056	conserved hypotet
15	43	44.8	370	2 H70423	oxygen-independent
16	43	44.8	2104	2 D91286	hypothetical prote
17	43	44.8	2104	2 H86127	hypothetical prote
18	42.5	44.3	602	2 AB3542	GTP-binding protei
19	42	43.8	115	2 B86882	hypothetical prote
20	42	43.8	143	1 HMKM41	heat shock protein
21	42	43.8	547	2 A40656	hypothetical prote
22	42	43.8	547	2 AE1022	probable membrane
23	42	43.8	646	2 D81674	conserved hypotet
24	42	43.8	1035	2 S18512	cell division cont
25	42	43.8	1250	2 A39578	SSDI protein - yea
26	41	42.7	141	2 E64368	hypothetical prote
27	41	42.7	155	2 C34965	hypothetical 17K p
28	41	42.7	155	2 S15576	lipi protein - Shi
29	41	42.7	231	2 B98297	hypothetical prote

30	41	42.7	231	2 AE2986	conserved hypothet
31	41	42.7	313	2 D95866	probable transcript
32	41	42.7	366	2 TDMSM4	monocyte surface g
33	41	42.7	518	1 S4183	thiamin-phosphate
34	41	42.7	605	2 D83007	regulatory protein
35	41	42.7	621	2 A71516	hypothetical prote
36	41	42.7	687	2 T09051	pepa protein - pse
37	41	42.7	708	2 JC6329	yeast secretory pr
38	41	42.7	1036	2 F82263	probable multidrug
39	41	42.7	1338	2 T40993	protein kinase cck
40	41	42.7	1652	2 I50711	complement C3 prec
41	40	41.7	143	1 HKM48	heat shock protein
42	40	41.7	208	2 A70122	glucose inhibited
43	40	41.7	247	2 E90656	hypothetical prote
44	40	41.7	247	2 F85507	hypothetical prote
45	40	41.7	332	2 JC4535	urate oxidase (EC
46	40	41.7	373	2 B90596	cell division prot
47	40	41.7	429	2 H90157	aspartyl-tRNA synt
48	40	41.7	448	2 C82936	signal recognition
49	40	41.7	638	2 S51266	trsg protein - yer
50	40	41.7	739	2 S39975	stringent response
51	40	41.7	819	1 B72128	endopeptidase Ia (
52	40	41.7	819	2 E86494	lon ATP-dependent
53	40	41.7	900	2 F97585	alanyl-tRNA synth
54	40	41.7	1042	2 H70203	isoleucine--trna 1
55	40	41.7	1151	2 T04657	hypothetical prote
56	40	41.7	1228	2 S00873	paraportal crystal
57	40	41.7	1388	2 S70633	serine/threonine-s
58	40	41.7	1388	2 S74245	serine/threonine-s
59	40	41.7	1770	2 S56221	hypothetical prote
60	39.5	41.1	603	2 AE2119	GTP-binding elonga

## ALIGNMENTS

RESULT 1  
GNNYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A: coat protein 1B: coat protein 1C: coat protein 1D: core  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Olsen, J.I., Ricehurst, J.R., Purcell, R.H., Buckler-White, A., Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1492/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: core protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1778/Product: protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0% Score 96: DB 1: Length 2227;  
Best Local Similarity 100.0% Pred. No. 2.8e-07;  
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0;  
OY 1 KINLADRMGLSGVQEIKEQ 20

Db 961 KINLADRMGLSGVQETKEQ 980

## RESULT 2

GNMNR  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; coat protein 3A; coat protein 3B; coat protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Majorian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nesselrooy, Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MUID:85190549  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NA>  
C:Cross-references: GB:K02990; NID:g329596; PIDN:AAA5472.1; PID:g329597  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 961 KINLADRMGLSGVQETKEQ 980

RESULT 3

GNMNR  
genome polyprotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; coat protein 3A; coat protein 3B; coat protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.I.; Rosenblum, B.; Titchhurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, F.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with  
A:Reference number: A94149; MUID:87115701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <CO>  
C:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA5471.1; PID:g329595  
A:Note: submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>

F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 961 KINLADRMGLSGVQETKEQ 980

## RESULT 4

GNMNR  
genome polyprotein - human hepatitis A virus (strain MB) (MB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; core protein 2B; core protein 2C; coat protein 3A; coat protein 3B; coat protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Kiehn, R.; Wimmer, E.; Deinhard  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso  
A:Reference number: J50303; MUID:88045071  
A:Accession: J50303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
C:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h  
F:1-245/Product: coat protein 1A #status predicted <VP4>  
F:246-491/Product: coat protein 1B #status predicted <VP2>  
F:492-836/Product: coat protein 1C #status predicted <VP3>  
F:837-980/Product: core protein 2A #status predicted <VP1>  
F:981-1108/Product: core protein 2B #status predicted <VP2B>  
F:1109-1438/Product: core protein 2C #status predicted <VP2C>  
F:1439-1496/Product: protein 3A #status predicted <VP3A>  
F:1497-1519/Product: genome-linked protein VPg #status predicted <VPg>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 961 KINLADRMGLSGVQETKEQ 980

## RESULT 5

GNMNR  
genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; core protein 2B; core protein 2C; coat protein 3A; coat protein 3B; coat protein 3D  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tsaev, S.A.  
Submitted to JTPD, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BA00766.1; PID:g222598  
R:Tsaev, S.A.; Emerson, S.U.; Balayan, M.S.; Titchhurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure  
A:Reference number: J01080; MUID:91311420  
A:Contents: annotation



A>Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdllov, E.D.; Chizhik  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL>  
A:Cross-references: EMBL:X15461; NID:g61971; PIDN:CAA33490.1; PID:g930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdllov, E.D.; Chizhik  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1436/Product: core protein 2C #status predicted <C2C>  
F:1437-1458/Product: protein 3A #status predicted <P3A>  
F:1459-1531/Product: protein 3B #status predicted <P3B>  
F:1532-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 100.0%; Score 96; DB 1; Length 2230;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKEQ 20  
DB 965 KINLADRLGLSGVQEIKEQ 984

RESULT 6  
B25199  
heat shock 16k protein 2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 13-Aug-1999  
C:Accession: B25199  
R:Jones, D.; Rusanak, R.H.; Kay, R.J.; Candido, E.P.M.  
J. Biol. Chem. 261, 12006-12015, 1986  
A:Title: Structure, expression, and evolution of a heat shock gene locus in Caenorhabdit  
A:Reference number: A92555; MUID:86304344  
A:Accession: B25199  
A:Molecule type: DNA  
A:Residues: 1-145 <JON>  
A:Cross-references: GB:M14334; NID:g156338; PIDN:AAA28071.1; PID:g156340  
C:Superfamily: alpha-crystallin

Query Match 49.0%; Score 47; DB 2; Length 145;  
Best Local Similarity 55.6%; Pred. No. 2.2;  
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKE 18  
DB 65 KINLADRLGLSGVQEIKE 82

RESULT 7  
S33416  
heat shock protein hsp20 - nematode (Nippostrongylus brasiliensis)  
C:Species: Nippostrongylus brasiliensis  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 13-Aug-1999  
C:Accession: S33416  
R:Twedde, S.; Gttrg, M.E.; Ingram, L.; Selkirk, M.E.

submitted to the EMBL Data Library, April 1993  
A:Description: The expression of a small heat shock homologue is developmentally regu  
A:Reference number: S33416  
A:Accession: S33416  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-172 <TWE>  
A:Cross-references: EMBL:X71663; NID:g297865; PIDN:CAA50655.1; PID:g297866  
C:Superfamily: alpha-crystallin

Query Match 49.0%; Score 47; DB 2; Length 172;  
Best Local Similarity 45.0%; Pred. No. 2.6;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKEQ 20  
DB 80 KVQLDDRLTVEGMQEVNTE 99

RESULT 8  
T06303  
eno11-CoA hydratase homolog F11C18.10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
C:Accession: T06303  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysmaert, C.; Dasseville, R.; De Clerck, R.;  
ewes, H.W.; Meyer, K.F.X.; Scheller, C.  
submitted to the Protein Sequence Database, April 1999  
A:Reference number: Z15589  
A:Accession: T06303  
A:Molecule type: DNA  
A:Residues: 1-414 <BEV>  
A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.10  
A:Experimental source: cultivar Columbia; BAC clone F11C18  
C:Genetics:  
A:Gene: ATSP:F11C18.10  
A:Map position: 4  
A:Introns: 42/3; 75/2; 108/1; 134/3; 162/3; 191/1; 219/3; 260/2; 280/3; 313/3; 347/3;

Query Match 45.8%; Score 44; DB 2; Length 414;  
Best Local Similarity 52.6%; Pred. No. 22;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKE 19  
DB 260 KIELIDRYFGIDTVEEIE 278

RESULT 9  
S72609  
GTP-binding membrane protein lepa - Mycobacterium leprae  
N:Alternate names: protein B1937\_F3\_81  
C:Species: Mycobacterium leprae  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 02-Feb-2001  
C:Accession: S72609  
R:Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, November 1993  
A:Reference number: S72609  
A:Accession: S72609  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-646 <SMI>  
A:Cross-references: EMBL:U00016; NID:g466961; PIDN:AAA17177.1; PID:g466991  
C:Genetics:  
A:Gene: lepa  
A:Start codon: GTG  
C:Superfamily: GTP-binding membrane protein lepa; translation elongation factor Tu ho  
C:Keywords: GTP binding; membrane protein; nucleotide binding; P-loop  
F:45-177/Domain: translation elongation factor Tu homology <TUN>  
F:51-58/Region: nucleotide-binding motif A (P-loop)  
F:174-177/Region: GTP-binding NKXD motif

Query Match 45.8%; Score 44; DB 2; Length 646;  
 Best Local Similarity 71.4%; Pred. No. 35;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGV 14  
 | | | | | | | | | |  
 Db 57 KSTLADRMGLQTCV 70

## RESULT 10

G70683  
 probable lepa - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 02-Feb-2001

C:Accession: G70683  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; M01D:98255987

A:Accession: G70683

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-653 <COL>

A:Cross-references: GB:281368; GB:ALJ23456; NID:g3261656; PIDN:CAB03723.1; PID:g1655655

A:Experimental source: strain H37RV

C:Genetics: lepa

A:Superfamily: GTP-binding membrane protein lepa; translation elongation factor Tu homol

C:Keywords: GTP binding; nucleotide binding; P-loop

F:53-181/Domain: translation elongation factor Tu homology <ETD>

F:59-66/Region: nucleotide-binding motif A (P-loop)

F:178-181/Region: GTP-binding NKXD motif

Query Match 45.8%; Score 44; DB 2; Length 653;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGV 14  
 | | | | | | | | | |  
 Db 65 KSTLADRMGLQTCV 78

## RESULT 11

S67595  
 hypothetical protein YDL060W - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2544

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Jun-2000

C:Accession: S67595

R:Bioecker, H.; Brandt, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67587

A:Accession: S67595

A:Molecule type: DNA

A:Residues: 1-788 <BLO>

A:Cross-references: EMBL:Z74108; NID:g1431062; PID:g1431063; GSPDB:GN00004; MIPS:YDL060W

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YDL060W

A:Map position: 4L

Query Match 45.8%; Score 44; DB 2; Length 798;  
 Best Local Similarity 61.5%; Pred. No. 44;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 8 MGLSGVQETKEQ 20

Db 159 VFGLSGVQEWDEE 171  
 : | | | | | | | | | |

## RESULT 12

B24289  
 heat shock protein 16-1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 28-Sep-1987 #sequence\_revision 30-Sep-1993 #text\_change 29-Oct-1999

C:Accession: B24289; S31037; T25927; T25930

R:Rusnak, R.H.; Candido, E.P.M.

Mol. Cell. Biol. 5, 1268-1278, 1985

A:Reference number: A24289; M01D:85295957

A:Accession: B24289

A:Molecule type: DNA

A:Residues: 1-145 <RUS>

A:Cross-references: EMBL:R03273; NID:g156333; PIDN:AAA28068.1; PID:g156336

A>Note: the author translated the codon GAT for residue 17 as Tyr

R:Ray, R.U.; Rusnak, R.H.; Jones, D.; Mathias, C.; Candido, E.P.M.

Nucleic Acids Res. 15, 3723-3741, 1987

A:Title: Expression of intron-containing C. elegans heat shock genes in mouse cells d

y effect of heat shock on the mammalian splicing apparatus.

A:Reference number: S31036; M01D:87231065

A:Accession: S31037

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 41-63 <RAY>

A:Cross-references: EMBL:M31340; NID:g156343; PIDN:AAA28073.1; PID:g552068

R:Bradshaw, H.

submitted to the EMBL Data Library, July 1996

A:Description: The sequence of C. elegans cosmid T27E4.

A:Reference number: Z20111

A:Accession: T25927

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-145 <RAY>

A:Cross-references: EMBL:U64837; PIDN:AA04839.1; GSPDB:GN00023; CESP:hsp-16A

A:Experimental source: strain Bristol N2; clone T27E4

A:Accession: T25930

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-145 <BR2>

A:Cross-references: EMBL:U64837; PIDN:AA04842.1; GSPDB:GN00023; CESP:hsp-16A

A:Experimental source: strain Bristol N2; clone T27E4

C:Genetics: CESP:hsp-16A

A:Map position: 5

A:Insertions: 42/3

C:Superfamily: alpha-crystallin

Query Match 44.8%; Score 43; DB 2; Length 145;  
 Best Local Similarity 45.0%; Pred. No. 10;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVQETKEQ 20  
 | | | | | | | | | |  
 Db 65 KINLADRMGLSGVQETKEQ 84

## RESULT 13

S64321  
 hypothetical protein YGR030c - yeast (Saccharomyces cerevisiae).

N:Alternate names: hypothetical protein G4068

C:Species: Saccharomyces cerevisiae

C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 29-Oct-1999

C:Accession: S64321

R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64071

A:Accession: S64321

A:Molecule type: DNA

A:Residues: 1-158 <RIE>

A:Cross-references: EMBL:272815; NID:g1323007; PID:e243932; PID:g1323008; GSPDB:GN00007;  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YGR030C  
A:Map position: 7R

Query Match 44.8%; Score 43; DB 2; Length 158;  
Best Local Similarity 64.3%; Pred. No. 11;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGV 14  
DB 65 QINMADRSGLGQV 78

RESULT 14  
AC0056  
Conserved hypothetical protein YPO0454 [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AC0056

R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MWID:21470413; PMID:11586360

A:Accession: AC0056

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 <RUR>

A:Cross-references: GB:AL590842; PIDN:CAC89310.1; PID:g15978546; GSPDB:GN00175

C:Genetics:

A:Gene: YPO0454

C:Superfamily: Escherichia coli conserved yjx protein

Query Match 44.8%; Score 43; DB 2; Length 180;  
Best Local Similarity 47.1%; Pred. No. 13;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 LADRMGLSGVOEIKQ 20  
DB 120 LQDEMAVLGSGISNVKQ 136

RESULT 15  
H70423  
Oxygen-independent coproporphyrinogen III oxidase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 11-Jun-1999

C:Accession: H70423

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MWID:98196666

A:Accession: H70423

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-370 <NOF>

A:Cross-references: GB:AE000739; NID:g2983813; PIDN:AAC07371.1; PID:g2983815; GB:AE00065

A:Experimental source: strain VFS

C:Genetics:

A:Gene: hemF

C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 44.8%; Score 43; DB 2; Length 370;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 6 DRMLGSLGVOEI 17  
DB 78 DKVFGSLGVKEI 89

RESULT 16

D91286

hypothetical protein ECs5260 [Imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D91286

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MWID:21156231; PMID:11258796

A:Accession: D91286

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2104 <HAY>

A:Cross-references: GB:BA000007; PIDN:BAB38683.1; PID:g13364738; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs5260

Query Match 44.8%; Score 43; DB 2; Length 2104;  
Best Local Similarity 47.8%; Pred. No. 1.9e+02;  
Matches 11; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

OY 2 INLADRMGLSGVOEIKQ 20  
DB 775 VDLADRRMRGSPFKELSGDLTRKQ 797

RESULT 17

H86127

hypothetical protein Z5898 [Imported] - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: H86127

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; DiMaio, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MWID:21074935; PMID:11206551

A:Accession: H86127

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2104 <STO>

A:Cross-references: GB:AE005174; NID:g12519303; PIDN:AAG59484.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z5898

Query Match 44.8%; Score 43; DB 2; Length 2104;  
Best Local Similarity 47.8%; Pred. No. 1.9e+02;  
Matches 11; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

OY 2 INLADRMGLSGVOEIKQ 20  
DB 775 VDLADRRMRGSPFKELSGDLTRKQ 797

RESULT 18

AB3542

gtp-binding protein lepa [Imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002

C:Accession: AB3542

R:DelVecchio, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov  
., Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AB3342  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-602 <RUB>  
A:Cross-references: GB:AE008918; PIDN:JAL53501.1; PID:g17984405; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10260  
A:Map position: II  
C:Superfamily: GRP-binding membrane protein lepa; translation elongation factor Tu homolog

Query Match	44.3%	Score 42.5;	DB 2;	Length 602;
Best Local Similarity	43.5%	Pred. No. 58;		
Matches 10;	Conservative 6;	Mismatches 4;	Indels 3;	Gaps 1;

OY 1 KINLADRMILGSLG---VOEIKEQ 20  
| | | | | : : : : : | : : : : : |  
db 21 KSTLADRLILQLGGLDPTREMKDQ 43

```

RESULT      19
B86882
hypothetical protein yveE [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86882
R:Boitlin, A.; Winkler, P.; Manger, S.; Tallon, O.; Melame, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:2123;186; PMID:11337471
A:Accession: B86882
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-115 <STO>
A:Cross-references: GB:AE00176; PID:q12725110; PIDN:AAK06156.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yveE

Oy      3  NLADRMILSGVQELKE 19
||:||||:|:|
DB      6  NIVDMRIGLEKREFEE 22

Query Match      43.8%, Score 42; DB 2; Length 115;
Best Local Similarity 47.1%; Pred. No. 12;
Matches      8; Conservative      3; Mismatches      6; Indels      0; Gaps      0;

```

```

RESULT 20
HHKMA1
heat shock protein 16-41 - Caenorhabditis elegans
N:Alternate names: heat shock protein 16 2
C:Species: Caenorhabditis elegans
C:Date: 25-Feb-1985 #sequence revision 19-Oct-1995 #text_change 22-Jun-1999
C:Accession: A25199; A38884; A02917
R:Jones, D.; Rusanak, R.H.; Kay, R.J.; Candido, E.P.M.
J. Biol. Chem. 261, 12006-12015, 1986
A:Title: Structure, expression, and evolution of a heat shock gene locus in Caenorhabditis elegans
A:Reference number: A92555; MUID:863A344
A:Accession: A25199
A:Molecule type: DNA
A:Residues: 1-143 <JON>
R:Candido, E.P.M.
submitted to Genbank, November 1985
A:Reference number: A38884
A:Accession: A38884
A:Molecule type: mRNA
A:Residues: 47-143 <CAN>

```

A:Cross-references: GB:X01577; NID:66758; PIDDN:CAA5732.1; PID:g780186  
R:Rusnak, R.H.; Jones, D.; Candido, E.P.M.  
Nucleic Acids Res. 11, 3187-3205, 1983  
A:Title: Cloning and analysis of cDNA sequences coding for two 16 kilodalton heat shock  
A:Reference number: A93467; MUID:83220736  
A:Accession: A02917  
A:Molecule type: mRNA  
A:Residues: 'KICSFV', 47-143 <RUS>  
A:Cross-references: GB:X01864; NID:g156331; PIDDN:AAA28065.1; PID:g156332  
A:Note: the authors translated the codon UUG for residue 46 as Phe  
A:Note: this sequence has been revised in reference A38884  
C:Superfamily: alpha-crystallin  
C:Keywords: heat shock; stress-induced protein

Query Match	43.8%	Score 42	DB 1	Length 143
Best Local Similarity	45.0%	Pred. No. 15		
Matches	9	Conservative	3	Mismatches
			8	Indels
			0	Gaps
Qy	1	KINLADRMIGLSGVOEIKQ	20	
db	69	KIKLDGRLEIKEGIOETKSE	88	

RESULT 21  
A40656  
hypothetical protein 1 (pmra 5' region) - *Salmonella typhimurium* (strain LT2)  
C:Species: *Salmonella typhimurium*  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 01-Dec-2000  
C:Accession: A40656  
R:Roland, K.L.; Martin, L.E.; Esther, C.R.; Spitznagel, J.K.  
J. Bacteriol. 175:4154-4164, 1993  
A:Title: Spontaneous pmra mutants of *Salmonella typhimurium* LT2 define a new two-compo  
A:Reference number: A40656; MUID:93308095  
A:Accession: A40656  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-547 <RDI>  
A:Cross-references: GB:J13395; NID:g295209; PIDN:AAA72364.1; PID:g295210  
C:Superfamily: *Escherichia coli* yidB protein  
:Keywords: transmembrane protein

RESULT 22  
AE1022  
probable membrane protein STY4492 [imported] - *Salmonella enterica* subsp. *enterica* serovar typhi  
C:Species: *Salmonella enterica* subsp. *enterica* serovar typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE1022  
R:Packhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerton, P.; Croplin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
S.; Moulé, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.;  
Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AE1022  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-547 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD09278.1; PID:g16505282; GSPDB:GN00176  
C:Gene: STY4492





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:10:12 ; Search time 59.43 Seconds  
(Without alignments)  
13.030 Million cell updates/sec

Title: US-09-171-432a-46

Perfect score: 96  
Sequence: 1 KINLADRMGLSGVQEIKEQ 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2226	POLG_HPAV2	P26560 hepatitis a
2	96	100.0	2226	POLG_HPAV4	P26561 hepatitis a
3	96	100.0	2226	POLG_HPAV8	P26562 hepatitis a
4	96	100.0	2227	POLG_HPAVH	P08617 hepatitis a
5	96	100.0	2227	POLG_HPAVL	P06441 hepatitis a
6	96	100.0	2227	POLG_HPAVM	P13901 hepatitis a
7	96	100.0	2230	POLG_HPAVS	P14553 simian hepa
8	48	50.0	622	LEPA_STRCO	Q9169 streptomyce
9	47	49.0	145	HS12_CAEEL	P06562 caenorhabd
10	47	49.0	172	HS20_NIPBR	Q07160 nippstrong
11	44	45.8	646	LEPA_MYCLE	P53530 mycobacteri
12	44	45.8	653	LEPA_MYCTU	P71739 mycobacteri
13	44	44.8	145	HS11_CAEEL	P34666 caenorhabd
14	43	44.8	158	POPE_YEAST	P53218 saccharomyc
15	42	43.8	143	HS16_CAEEL	P06561 caenorhabd
16	42	43.8	151	SODC_HALRO	P81926 halocynthia
17	42	43.8	547	YJDB_SALTY	P36555 salmonella
18	42	43.8	621	YJ08_CHLMU	O9P148 chlamydia m
19	42	43.8	1035	CC68_YEAST	P13258 saccharomyc
20	42	43.8	1250	SSDI_YEAST	P24276 saccharomyc
21	41	42.7	141	NIRK_METUA	Q57969 methanococc
22	41	42.7	155	IPPI_SHIFL	P18008 shigella fl
23	41	42.7	313	CBRI_RHIME	P58332 rhizobium m
24	41	42.7	313	CBRI_RHIME	P58332 rhizobium m
25	41	42.7	366	CDI4_MOUSE	P10810 mus musculu
26	41	42.7	518	THI4_SCHPO	P40366 s probabie
27	41	42.7	621	Y425_CHLTR	O84432 chlamydia t
28	41	42.7	1338	CEK1_SCHPO	P38938 schizosacch
29	40	41.7	143	HS17_CAEEL	P02513 caenorhabd
30	40	41.7	208	GIDB_BORBU	P53353 borrelia bu
31	40	41.7	316	OC03_HUMAN	O9UGF7 homo sapien
32	40	41.7	332	URIC_BACSB	O45657 bacillus sp
33	40	41.7	698	MCHE_ECOLI	O9EXN5 escherichia

34	40	41.7	739	1	RELA_STREQ	O54089 streptococc
35	40	41.7	819	1	LON_CHLNP	O92914 chlamydia p
36	40	41.7	1042	1	SVI_BORBU	O51773 borrelia bu
37	40	41.7	1227	1	C1BE_BACTU	O85805 bacillus th
38	40	41.7	1228	1	C1BA_BACTK	P05517 bacillus th
39	40	41.7	1229	1	C1BB_BACTM	O45739 bacillus th
40	40	41.7	1231	1	C1BC_BACTM	O92a25 bacillus th
41	40	41.7	1233	1	C1BD_BACTM	O45774 bacillus th
42	40	41.7	1770	1	R115_YEAST	P43565 saccharomyc
43	39	40.6	119	1	Y033_ARCEU	O30238 archaeglob
44	39	40.6	251	1	ASTA_ASTFL	P07564 astacus flu
45	39	40.6	365	1	AROC_HELPY	O92111 helicobacte
46	39	40.6	365	1	AROC_HELPY	P56122 helicobacte
47	39	40.6	454	1	SR54_AQUAE	O67615 aquifex aeo
48	39	40.6	473	1	ATPB_RICCN	O92968 rickettsia
49	39	40.6	482	1	HOXA_ALCEU	P29267 alcaligenes
50	39	40.6	748	1	PHY1_STAY3	O55168 synechocyst
51	39	40.6	819	1	LON_CHLMU	O9PX30 chlamydia m
52	39	40.6	819	1	LON_CHLMU	O84348 chlamydia t
53	38.5	40.1	269	1	MIND_GUTTH	O78436 guillardi
54	38.5	40.1	266	1	MIND_MESVI	O9MUN5 mesostigma
55	38.5	40.1	745	1	K6A6_HUMAN	O9UK32 homo sapien
56	38	39.6	138	1	SODC_LAMCR	P81036 lampyricus
57	38	39.6	152	1	ENRN_BPT3	P20314 bacterioph
58	38	39.6	154	1	SODC_BRARE	O73872 brachydanio
59	38	39.6	185	1	TD52_MOUSE	O62393 mus musculu
60	38	39.6	225	1	RS5_METVA	P14036 methanococc

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	2226 AA.
ID	POLG_HPAV2			
AC	P26560;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins			
DE	P2A TO P2C; Probable proteins P2A TO P2C; RNA-directed RNA polymerase			
DE	P3D (EC 2.7.7.48)].			
OS	Hepatitis A virus (strain 24a).			
OC	Hepatitis A virus (strain 24a).			
OC	Hepatitis A virus (strain 24a).			
OC	Hepatitis A virus (strain 24a).			
OX	NCBI_TaxID=12094;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9162758; PubMed=1705995;			
RA	Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,			
RA	Cromean T., Jansen R.W.;			
RT	"Antigenic and genetic variation in cytopathic hepatitis A virus			
RT	variants arising during persistent infection: evidence for genetic			
RT	recombination."			
RL	J. Virol. 65:2056-2065(1991).			
CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,			
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,			
CC	VP3, AND VP4.			
CC	-1- PMW: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	OR SEND AN EMAIL TO <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: M59810; AAA45468.1; -			
DR	MEROPS: C03.005;			
DR	InterPro: IPR000605; RNA_helicase.			
DR	InterPro: IPR001205; RNA_pol_P3D.			

DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;  
  
Query Match 100.0%; Score 96; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVQEIKEQ 20  
DB 961 KINLADRMGLSGVQEIKEQ 980  
|||||  
POLG\_HPAV4 STANDARD; PRT; 2226 AA.  
ID POLG\_HPAV4  
AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain 43c).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Sheldale P.A., Ping L.H., Feinstone S.M.,  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M59809; AAA45469.1; -  
CC  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;  
  
Query Match 100.0%; Score 96; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVQEIKEQ 20  
DB 961 KINLADRMGLSGVQEIKEQ 980  
|||||  
POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
ID POLG\_HPAV8  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain 18f).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Sheldale P.A., Ping L.H., Feinstone S.M.,  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M59808; AAA45467.1; -  
CC  
DR MEROPS: C03.005; -  
DR InterPro: IPR000605; RNA\_helicase.  
DR InterPro: IPR001205; RNA\_pol\_P3D.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;



FT CHAIN 1088 1422 CORE PROTEIN P2C.  
 FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
 FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
 FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
 FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
 SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8DB8 CRC64;

Query Match 100.0%; Score 96; DB 1; Length 2226;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
 DB 961 KINLADRMGLSGVOEIKQ 980

RESULT 4  
 POLG\_HPAVL STANDARD; PRT; 2227 AA.  
 AC P08617; P08643; Q81082;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain HM-175).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.  
 NC NCB1\_TaxID=12098;  
 RX MEDLINE=87061253; PubMed=3023706;  
 RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A., Baroudy B.M.;  
 RT "Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses.";  
 RT J. Virol. 61:50-59(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATTENUATED;  
 RX MEDLINE=87175701; PubMed=3031686;  
 RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Felstone S.M., Purcell R.H.;  
 RT "Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with wild-type virus.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85166289; PubMed=2984684;  
 RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr., Purcell R.H., Felstone S.M.;  
 RT "Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA polymerase.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
 CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT SHOWN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M14114; AAA45475.1; -;  
 CC EMBL: M14707; AAA45465.1; -;  
 DR EMBL: M14707; AAA45466.1; ALT\_INIT.  
 DR EMBL: M16632; AAA45471.1; -;  
 DR PIR: A25981; GNNYHM.  
 DR PIR: A25914; GNNYMK.  
 DR PIR: A03905; A03905.  
 DR MEROPS: C03.005; -;  
 DR InterPro: IPR000605; RNA\_helicase.  
 DR InterPro: IPR001205; RNA\_pol\_P3D.  
 DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam: PF00910; RNA\_helicase; 1.  
 KW Polypeptide; Coat protein; Core protein; Core protein: Transferase; RNA-directed RNA polymerase; Hydrolyase; Thiol protease.  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 836  
 FT CHAIN 837 980  
 FT CHAIN 981 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1496  
 FT CHAIN 1497 1519  
 FT CHAIN 1520 1738  
 FT CHAIN 1739 2227  
 FT VARIANT 77 77  
 FT VARIANT 764 764  
 FT VARIANT 821 821  
 FT VARIANT 1052 1052  
 FT VARIANT 1062 1062  
 FT VARIANT 1118 1118  
 FT VARIANT 1151 1151  
 FT VARIANT 1163 1163  
 FT VARIANT 1277 1277  
 FT VARIANT 1500 1500  
 FT VARIANT 1805 1805  
 FT VARIANT 1930 1930  
 SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;  
 K -> R (IN ATTENUATED STRAIN).  
 E -> V (IN ATTENUATED STRAIN).  
 N -> S (IN ATTENUATED STRAIN).  
 A -> V (IN ATTENUATED STRAIN).  
 G -> A (IN ATTENUATED STRAIN).  
 K -> M (IN ATTENUATED STRAIN).  
 E -> K (IN ATTENUATED STRAIN).  
 F -> S (IN ATTENUATED STRAIN).  
 V -> I (IN ATTENUATED STRAIN).  
 H -> Y (IN ATTENUATED STRAIN).  
 D -> N (IN ATTENUATED STRAIN).  
 S -> T (IN ATTENUATED STRAIN).

Query Match 100.0%; Score 96; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
 DB 961 KINLADRMGLSGVOEIKQ 980

RESULT 5  
 POLG\_HPAVL STANDARD; PRT; 2227 AA.  
 AC P06441;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polypeptide [contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain LA).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.  
 NC NCB1\_TaxID=12099;  
 RX MEDLINE=85190549; PubMed=2986127;  
 RA Meritarian R., Caput D., Gee W.W., Potter S.J., Renard A., Meriweather J., van Nest G., Dina D.;  
 RT "Primary structure and gene organization of human hepatitis A virus.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).

CC -1- SUBUNIT. THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: K02990; AAA5472.1; -  
CC PIR: A03903; GNNYHR.  
CC MEROPS: C03.005; -  
CC InterPro: IPR001205; RNA\_pol\_P3D.  
CC Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
CC DR InterPro: IPR001205; RNA\_pol\_P3D.  
CC Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
CC DR Polyprotein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
CC KW CHAIN 1 23  
CC FT CHAIN 24 245  
CC FT CHAIN 246 491  
CC FT CHAIN 492 836  
CC FT CHAIN 837 980  
CC FT CHAIN 981 1076  
CC FT CHAIN 1077 1422  
CC FT CHAIN 1423 1484  
CC FT CHAIN 1485 1507  
CC FT CHAIN 1508 1678  
CC FT CHAIN 1679 2227  
CC SO SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;  
  
Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 961 KINLADRMGLSGVQEIKEQ 980  
  
RESULT 6  
POLG\_HPAVM STANDARD; PRT; 2227 AA.  
ID POLG\_HPAVM Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4, Core proteins  
DE P2A TO P2C, Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain MBB).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=8045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wessel T., Klehn R., Wimmer E.,  
RA Delhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
RT virus (isolate MBB).";  
RL VIRUS REF. 8:153-171(1987).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M20273; AAA5474.1; -  
CC PIR: J50303; GNNYHB.  
CC MEROPS: C03.005; -  
CC InterPro: IPR001205; RNA\_pol\_P3D.  
CC Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
CC DR Polyprotein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
CC KW CHAIN 1 23  
CC FT CHAIN 24 245  
CC FT CHAIN 246 491  
CC FT CHAIN 492 836  
CC FT CHAIN 837 980  
CC FT CHAIN 981 1087  
CC FT CHAIN 1088 1422  
CC FT CHAIN 1423 1496  
CC FT CHAIN 1497 1519  
CC FT CHAIN 1520 1738  
CC FT CHAIN 1739 2227  
CC SO SEQUENCE 2227 AA; 251425 MW; EC9835BD2A7C86349 CRC64;  
  
Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 961 KINLADRMGLSGVQEIKEQ 980  
  
RESULT 7  
POLG\_HPAVS STANDARD; PRT; 2230 AA.  
ID POLG\_HPAVS P14553;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4, Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Simian hepatitis A virus (strain AGM-27).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=91311420; PubMed=1649901;  
RA Tsarev S.A., Emerson S.U., Balayan M.S., Titchhurst J.R.,  
RA Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
RT structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RA MEDLINE=89232168; PubMed=2541023;  
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,  
RA Sverdlov E.D., Chizhikov V.E., Bilnov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
RT and simian hepatitis A viruses.";  
RL FEBS Lett. 247:425-428(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

---

```

DR EMBL: AL136503; CAB66240.1; -.
DR InterPro: IPR000795; GTP_EFTU.
DR InterPro: IPR004161; GTP_EFTU_D2.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF03144; GTP_EFTU_D2; 1.
DR PRINTS: PR00315; ELONGATNCT.
DR PROSITE: PS00301; EFACOR_GTP; 1.
KW GTP-binding.
FT NP_BIND 26 33 GTP (BY SIMILARITY).
FT NP_BIND 94 98 GTP (BY SIMILARITY).
FT NP_BIND 148 151 GTP (BY SIMILARITY).
SQ SEQUENCE 622 AA; 68378 MW; 83F5C76FA2A80C7C CRC64;
```

---

```

OY 1 KINLADRMGLSGVOEIK 19
   | ||||| :|| :| ::
Db 32 KSTLADRMGLTGVVEORQ 50

RESULT 9
ID HS12_CAEEL STANDARD; PRT; 145 AA.
AC P06582;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Heat shock protein HSP16-2.
GN HSP16-2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66304344; PubMed=3017958;
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
RT Caenorhabditis elegans that is flanked by repetitive elements.";
RL J. Biol. Chem. 261:12006-12015(1986).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

---

```

DR EMBL: M14334; AAA28071.1; -.
DR PIR: B25199;
DR InterPro: IPR002068; Crystallin_HSP20.
DR Pfam: PF00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 145 AA; 16242 MW; 8A73449P9161889 CRC64;
```

---

```

Query Match 49.0%; Score 47; DB 1; Length 145;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
```

---

```

OY 1 KINLADRMGLSGVOEIK 18
   ||||| :| :| :|||

```



RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CDC 1551 / Oshkosh;  
RA Pleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeJoy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
CC LEPA SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z81368; CAB03723.1; -;  
DR EMBL: AE007086; AAK46772.1; -;  
DR HSSP: P13551; 2EFG.  
DR TIGR: MT2476; -;  
DR TubercuList; Rv2404c; -;  
DR InterPro: IPR000795; GTP\_EFTU.  
DR InterPro: IPR004161; GTP\_EFTU\_D2.  
DR Pfam: PF00009; GTP\_EFTU; 1.  
DR Pfam: PF03144; GTP\_EFTU\_D2; 1.  
DR PROSITE: PS00301; EFATOR\_GTP; 1.  
DR GTP-binding; Complete proteome.  
DR NP\_BIND 59 66 GTP (BY SIMILARITY).  
DR NP\_BIND 124 128 GTP (BY SIMILARITY).  
DR NP\_BIND 178 181 GTP (BY SIMILARITY).  
FT NP\_BIND 178 181 GTP (BY SIMILARITY).  
SQ SEQUENCE 653 AA; 72395 MW; DA4AFE10B6C25755 CRC64;

Query Match 45.8%; Score 44; DB 1; Length 653;  
Best Local Similarity 71.4%; Pred. No. 15;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KINLADRLGLSGV 14  
| | | | | | | | | | | | | | | | | |  
Db 65 KSTLADRLGLTGV 78

RESULT 13  
HSL1\_CAEEL STANDARD; PRT; 145 AA.  
AC P34696;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DE Heat shock protein HSP16-1.  
GN (HSP16-1A OR T27E4.2) AND (HSP16-1B OR T27E4.8).  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodermidae; Caenorhabditis.  
OX NCBI\_Taxid=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85295957; PubMed=4033652;  
RA Russnak R.H., Candido E.P.M.;  
RT "Locus encoding a family of small heat shock genes in Caenorhabditis  
RT elegans: two genes duplicated to form a 3.8-kilobase inverted  
RT repeat.";

RL MOL. Cell. Biol. 5:1268-1278(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Bradshaw H.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 41-63 FROM N.A.  
RX MEDLINE=87231065; PubMed=3588308;  
RA Kay R.J., Russnak R.H., Jones D., Mathias C., Candido E.P.M.;  
RT "Expression of intron-containing C. elegans heat shock genes in mouse  
RT cells demonstrates divergence of 3' splice site recognition sequences  
RT between nematodes and vertebrates, and an inhibitory effect of heat  
RT shock on the mammalian splicing apparatus.";  
RL Nucleic Acids Res. 15:3723-3741(1987).  
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: K03273; AAA28068.1; -;  
DR EMBL: M31340; AAA28073.1; -;  
DR EMBL: U64837; AAB04842.1; -;  
DR EMBL: U64837; AAB04839.1; -;  
DR PIR: B24289; B24289.  
DR WormPep; T27E4.2; CE14249.  
DR WormPep; T27E4.8; CE14249.  
DR InterPro: IPR002068; Crystallin\_HSP20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
DR Heat shock; Multigene family.  
KW SEQUENCE 145 AA; 16253 MW; 06C36A1F06D15A11 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 145;  
Best Local Similarity 45.0%; Pred. No. 4.3;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 KINLADRLGLSGVQIEKQ 20  
| | | | | | | | | | | | | | | | | |  
Db 65 KINLADRLGLSGVQIEKTE 84

RESULT 14  
POP6\_YEAST STANDARD; PRT; 158 AA.  
AC P53218;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ribonucleases P/MRP protein subunit POP6 (EC 3.1.26.5) (RNases P/MRP  
DE 18.2 kDa subunit) (RNA processing protein POP6).  
GN POP6 OR YGR030C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE=97435481; PubMed=9290212;  
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;  
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae  
RT chromosome VII.";  
RL Yeast 13:1077-1090(1997).  
CC -1- FUNCTION: COMPONENT OF RIBONUCLEASE P, A PROTEIN COMPLEX THAT  
CC GENERATES MATURE tRNA MOLECULES BY CLEAVING THEIR 5' ENDS.

CC ALSO A COMPONENT OF RNASE MRP.  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-  
CC extra-nucleotide from tRNA precursor.  
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RNASE P AND RNASE MRP RNASE P  
CC RIBONUCLEOPROTEINS. RNASE P CONSISTS OF A RNA MOLEKY AND AT LEAST  
CC 8 PROTEIN SUBUNITS: POP1, POP3, POP4, POP5, POP6, POP7, POP8 AND  
CC RPL1.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z72815; CAA97018.1; -.  
DR SCD: S0003262; POP6.  
DR HYDROLASE; Nuclease protein; tRNA processing; Colled coll.  
KW DOMAIN 51 71 COILED COIL (POTENTIAL).  
SQ SEQUENCE 158 AA; 18210 MW; 6C27A73FAD521181 CRC64;  
Query Match 44.8%; Score 43; DB 1; Length 158;  
Best Local Similarity 64.3%; Pred. No. 4.7;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KINLADRMGLSGV 14  
Db 65 QINMADRSLGIGV 78  
RESULT 15  
HS16\_CAEEL STANDARD; PRT; 143 AA.  
ID HS16\_CAEEL STANDARD; PRT; 143 AA.  
AC P06581; P02514;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Heat shock protein HSP16-41.  
GN HSP16-41.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86304344; PubMed=3017958;  
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;  
RT "Structure, expression, and evolution of a heat shock gene locus in  
RT Caenorhabditis elegans that is flanked by repetitive elements.";  
RL J. Biol. Chem. 261:12006-12015(1986).  
RN [2]  
RP SEQUENCE OF 47-143 FROM N.A.  
RX MEDLINE=83220736; PubMed=6190129;  
RA Russnak R.H., Jones D., Candido E.P.M.;  
RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton  
RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with  
RT the small hsps of Drosophila.";  
RL Nucleic Acids Res. 11:3187-3205(1983).  
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
CC FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M14334; AAA28070.1; ALT\_SHP.

DR EMBL: X01577; CAA25732.1; -.  
DR PIR: A38884; HHRW41.  
DR PIR: A25199; A25199.  
DR InterPro: IPR002068; Crystallin\_HSP20.  
DR Pfam: PF00011; HSP20; 1.  
DR PROSITE: PS01031; HSP20; 1.  
KW Heat shock; Multigene family.  
SQ SEQUENCE 143 AA; 16252 MW; C1D0F59D26E36C24 CRC64;  
Query Match 43.8%; Score 42; DB 1; Length 143;  
Best Local Similarity 45.0%; Pred. No. 6.2;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
QY 1 KINLADRMGLSGVQEIKEQ 20  
Db 69 KIKIDGRLEKIEGIEYKSE 88

RESULT 16  
ID SODC\_HALRO STANDARD; PRT; 151 AA.  
AC P81926;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyuridae; Halocynthia.  
OX NCBI\_TaxID=7729;  
RN [1]  
RP SEQUENCE, AND CHARACTERIZATION.  
RC TISSUE=Hemocyte, and Plasma;  
RX MEDLINE=99302489; PubMed=10374259;  
RA Abe Y., Ishikawa G., Satch H., Azumi K., Yokosawa H.;  
RT "Primary structure and function of superoxide dismutase from the  
RT ascidian Halocynthia roretzi.";  
RL Comp. Biochem. Physiol. 122B:321-326(1999).  
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS. THE PLASMA SUPEROXIDE  
CC DISMUTASE HAS PHAGOCYTOSIS-STIMULATING ACTIVITY AND MAY PLAY AN  
CC IMPORTANT ROLE IN THE BIOLOGICAL DEFENSES OF THE ORGANISM.  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc (by similarity).  
CC -1- ENZYME REGULATION: INHIBITED BY KCN AND DIETHYLDITHIOCARBAMATE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
CC -1- PTM: ACETYLATED N-TERMINAL SERINE.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
DR HSSP: P15107; 1XSO.  
DR InterPro: IPR001424; SOD\_CU\_ZN.  
DR Pfam: PF00080; sdcu; 1.  
DR PRINTS: PR00068; CUZNDISMUTASE.  
DR PRODOM: PD000469; SOD\_CU\_ZN; 1.  
DR PROSITE: PS00087; SOD\_CU\_ZN\_1; 1.  
DR PROSITE: PS00332; SOD\_CU\_ZN\_2; FALSE\_NEG.  
KW Oxidoreductase; Copper; Zinc.  
FT METAL 43 43 COPPER (BY SIMILARITY).  
FT METAL 45 45 COPPER (BY SIMILARITY).  
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 68 68 ZINC (BY SIMILARITY).  
FT METAL 77 77 ZINC (BY SIMILARITY).  
FT METAL 80 80 ZINC (BY SIMILARITY).  
FT METAL 117 117 COPPER (BY SIMILARITY).  
FT DISULFID 54 143 BY SIMILARITY.  
SQ SEQUENCE 151 AA; 15489 MW; 162F181A82275AF0 CRC64;  
Query Match 43.8%; Score 42; DB 1; Length 151;  
Best Local Similarity 43.8%; Pred. No. 6.6;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 INLADRLGLSGVOEI 17  
 ID Y708\_CHLMU STANDARD; PRT; 621 AA.  
 DB 94 UNITDRMISLTGHERSI 109

RESULT 17

YDDB\_SALTY STANDARD; PRT; 547 AA.

AC P36555;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DE 01-MAR-2002 (Rel. 41, Last annotation update)

GN YDDB OR STM4293.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

NCBI\_TaxID=602;

RP SEQUENCE FROM N.A.

RC SRRAIN-LT2;

RA MEDLINE-93308095; PubMed-8391535;

RT Spontaneous pmra mutants of Salmonella typhimurium LT2 define a new

two-component regulatory system with a possible role in virulence.";

J. Bacteriol. 175:4154-4164(1993).

RL [2]

CC SEQUENCE FROM N.A.

RC SRRAIN-LT2 / SCSC1412 / ATCC 700720;

RA MEDLINE-21534948; PubMed-11677609;

RT McClelland M., Sanderson K.E., Spleth J., Clifton S.W., Latreille P.,

Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,

Leonard S., Nguyen C., Scott K., Holmes A., Grevell N., Mulvaney E.,

Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,

Waterston R., Wilson R.K.;

RA "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

Nature 413:852-856(2001).

CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

-I- SIMILARITY: BELONGS TO THE UPF0141 FAMILY.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

RESULT 18  
 ID Y708\_CHLMU STANDARD; PRT; 621 AA.  
 AC 09PJMB;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN TC0708.  
 OS Chlamydia muridarum.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

NCBI\_TaxID=83560;

OX [1]

RP SEQUENCE FROM N.A.

RC SRRAIN-MOPN / N199;

RA MEDLINE-20150255; PubMed-10684935;

RT Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

White O., Hickey E.K., Peterson J., Ullrich T., Berry K., Bass S.,

Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,

RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia

pneumoniae AR39.";

RT Nucleic Acids Res. 28:1397-1406(2000).

CC -I- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0512/CT425/TC0708

FAMILY.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 43.88; Score 42; DB 1; Length 621;  
 Best Local Similarity 44.48; Pred. No. 30;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 NLADRLGLSGVOEI 20

ID Y708\_CHLMU STANDARD; PRT; 1035 AA.

AC P32536;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

GN Cell division control protein 68.

OS CDC68 OR SPT16 OR SSF1 OR YGL207W.

OC Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

NCBI\_TaxID=4932;

RP SEQUENCE FROM N.A.

RC MEDLINE-92017853; PubMed-1833637;

RT "CDC68, a yeast gene that affects regulation of cell proliferation

and transcription, encodes a protein with a highly acidic carboxyl

terminus.";

RT Mol. Cell. Biol. 11:5718-5726(1991).

RL [2]

CC SEQUENCE FROM N.A.

```

RX MEDLINE=92017852; PubMed=1922073;
RA Malone E.A., Clark C.D., Chiang A., Winston F.;
RT "Mutations in SPT16/CDC68 suppress cis- and trans-acting mutations
RL that affect promoter function in Saccharomyces cerevisiae.";
RN Mol. Cell. Biol. 11:5710-5717(1991).
[3]
RX SEQUENCE FROM N.A.
RA MEDLINE=97298309; PubMed=9153757;
RA Feuerstein M., Simeonova L., Souciet J.-L., Potier S.;
RT "Analysis of 21.7 kb DNA sequence from the left arm of chromosome VII
RL reveals 11 open reading frames: two correspond to new genes.";
RL Yeast 13:475-477(1997).
CC -1- FUNCTION: PLAYS A ROLE IN GENERAL TRANSCRIPTION. IT HAS POSITIVE
CC AND NEGATIVE EFFECTS ON GENE EXPRESSION. REQUIRED FOR THE
CC APPROPRIATE SYNTHESIS DURING HEAT SHOCK. REQUIRED FOR CONTINUED
CC EXPRESSION OF CYCLIN GENES THAT DETERMINE THE PASSAGE THROUGH
CC START DURING CELL CYCLE CONTROL. IT MAY ACT AS AN ACIDIC
CC ACTIVATOR. ITS FUNCTION IS PROBABLY RELATED TO OTHER SPT GENES.
CC -1- PPM: PHOSPHORYLATED BY CASEIN KINASE II (POSSIBLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73533; -; NOT_ANNOTATED_CDS.
DR EMBL; Z72729; CAA96920.1; -.
DR PIR; S18512; S18512.
DR SCD; S0003175; SPT16.
DR Cell division; Cell cycle; Transcription regulation; Phosphorylation.
FT DOMAIN 958 1021 ASP, GLU-RICH (ACIDIC).
SQ SEQUENCE 1035 AA; 118629 MW; 4F01C772E299E2E6 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 1035;
Best Local Similarity 47.4%; Pred. No. 53;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKR 19
Db 159 EFNVIDISLGSKYWEYKD 177

RESULT 20
SSDI_YEAST STANDARD; PRT; 1250 AA.
ID SSD1_YEAST
AC P24276;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SSD1 protein (SRK1 protein).
GN SSD1 OR SRK1 OR C1A1 OR R1D1 OR YDR293C OR D9819.4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91172202; PubMed=1848673;
RA Sutton A., Immanuel D., Arndt K.T.;
RT "The S174 protein phosphatase functions in late G1 for progression
RL into S phase.";
RL Mol. Cell. Biol. 11:2133-2148(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=91246207; PubMed=1645445;
RA Wilson R.B., Brenner A.A., White T.B., Engler M.J., Gaughran J.P.,
RA Tatchell K.;
RT "The Saccharomyces cerevisiae SHK1 gene, a suppressor of hoy1 and
RL hsl1, may be involved in protein phosphatase function.";

```

```

RL Mol. Cell. Biol. 11:3369-3373(1991).
RN [3]
RX SEQUENCE FROM N.A.
RA STRAIN=8288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Ridsen L., Riles L., Tatch A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DDJ databases.
CC -1- FUNCTION: CAN SUPPRESS THE LETHALITY DUE TO DELETION OF S174, AND
CC PARTIALLY THE DEFECTS DUE TO BCY1 DISRUPTION. IS IMPLICATED IN THE
CC CONTROL OF THE CELL CYCLE G1 PHASE.
CC -1- MISCELLANEOUS: SEVERAL ALLELES OF SSD1 EXIST IN DIFFERENT YEAST
CC STRAINS.
CC -----
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60318; AAA35047.1; -.
DR EMBL; M63004; AAA35089.1; -.
DR EMBL; U51031; AAB64469.1; -.
DR PIR; A40263; A40263.
DR PIR; A39578; A39578.
DR SCD; S0002701; SSD1.
DR InterPro; IPR001900; Ribonuclease_II.
DR Pfam; PF00773; RNB; 1.
DR PROSITE; PS01175; RIBONUCLEASE_II; 1.
KM Mitosis; Cell cycle.
FT DOMAIN 60 67 POLY-GLN.
FT DOMAIN 71 74 POLY-GLN.
FT DOMAIN 135 141 POLY-ASN.
FT DOMAIN 214 217 POLY-PRO.
FT DOMAIN 527 530 POLY-GLU.
SQ SEQUENCE 1250 AA; 139954 MW; 3D50FD6838EBCA514 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 1250;
Best Local Similarity 42.1%; Pred. No. 65;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 INLADRMGLSGVOEIKR 20
Db 871 LNIIDRLTGFVINEIKRK 889

RESULT 21
NIKR_METJA STANDARD; PRT; 141 AA.
ID NIKR_METJA
AC O57969;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative nickel responsive regulator.
GN M0549.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

```



RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,  
RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
RA Ullrich T.R., Kelley J.M., Peterson J.D., Sedow P.W., Hanna M.C.,  
RA Colton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1- FUNCTION: Transcriptional regulator (potential).  
CC -1- COFACTOR: Binds 1 nickel ion per subunit (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE COPG/NIKR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U67504: AAB98541.1: -  
DR TIGR: M0549: -  
DR InterPro: IPR002145: COPG\_HTH\_4.  
DR Pfam: PF01402: HTH\_4; 1.  
KM Hypothetical protein; Transcription regulation; DNA-binding; Nickel;  
KW Metal-binding; Complete proteome.  
FT METAL 80 80 NICKEL (POTENTIAL).  
FT METAL 91 91 NICKEL (POTENTIAL).  
FT METAL 93 93 NICKEL (POTENTIAL).  
FT METAL 99 99 NICKEL (POTENTIAL).  
SQ SEQUENCE 141 AA; 16132 MW; 5BE0C5F09D38B5BD CRC64;  
  
Query Match 42.7%; Score 41; DB 1; Length 141;  
Best Local Similarity 46.7%; Pred. No. 9;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
OY 4 LADRMGLSGVOEIKR 18  
DB 115 LTDRMLTKGVQYK 129  
  
RESULT 22  
IPPL\_SHIFL STANDARD; PRT: 155 AA.  
AC P18008;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Ippi protein.  
GN Ippi.  
OS Shigella flexneri, and  
OS Shigella dysenteriae.  
OG Plasmid 210 kb invasion pWR100, and Plasmid 230 kb pMYSH6000.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=623, 622;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.flexneri; STRAIN=M90T / SEROTYPE 5;  
RX MEDLINE=89057927; PubMed=3057506;  
RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;  
RT "Characterization of invasion plasmid antigen genes (ipabCD) from  
RT Shigella flexneri.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.flexneri; STRAIN=2A; PLASMID=230 kb pMYSH6000;  
RX MEDLINE=9001179; PubMed=2552264;  
RA Sasakawa C., Adler B., Tobe T., Okada N., Nagai S., Komatsu K.,

RA Yoshikawa M.;  
RT "Functional organization and nucleotide sequence of virulence  
RT region-2 on the large virulence plasmid in Shigella flexneri 2a.";  
RL Mol. Microbiol. 3:1191-1201(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.flexneri; STRAIN=M90T / SEROTYPE 5;  
RX PLASMID=210 kb invasion pWR100;  
RX MEDLINE=89200844; PubMed=3071655;  
RA Baudry B., Kaczorek M., Sansonetti P.J.;  
RT "Nucleotide sequence of the invasion plasmid antigen B and C genes  
RT (ipab and ipac) of Shigella flexneri.";  
RL Microb. Pathog. 4:345-357(1988).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.dysenteriae; STRAIN=CG97;  
RA Yao R., Palchaudhuri S.;  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO E.COLI YGEG AND YERSINIA PLASMIDS ICRH.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M86530: RAD15224.1: -  
DR EMBL: J04117: AAA26521.1: -  
DR EMBL: X15319: CAA33380.1: -  
DR EMBL: M34849: AAA98423.1: -  
DR EMBL: X60777: CAA43189.1: -  
DR PIR: A31265: A31265.  
DR PIR: S06202; S06202.  
DR PIR: C34965; C34965.  
DR PIR: S15576; S15576.  
KM Plasmid; Virulence.  
SQ SEQUENCE 155 AA; 17756 MW; F46E39A02B8ADEC7 CRC64;  
  
Query Match 42.7%; Score 41; DB 1; Length 155;  
Best Local Similarity 46.7%; Pred. No. 9.9;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
  
OY 6 DRMLGSGVOEIKR 20  
DB 71 DYIMGLAIVQIKR 85  
  
RESULT 23  
CBRL\_RHME STANDARD; PRT: 313 AA.  
AC P58332;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Rubisco operon transcriptional regulator.  
GN CBBR OR RB0196 OR SMB20203.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymb (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396508; PubMed=11481431;  
RA Firan T.M., Weidner S., Wong K., Buhrmester J., Chai P.,  
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
RA Golding B., Puenhler A.;  
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
RT fixing endosymbiont Sinorhizobium meliloti.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:9389-9894(2001).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE CBZ OPERON FOR  
 CC RUBISCO AND OTHER CALVIN CYCLE GENES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF211846; AAF25374.1; -  
 CC InterPro: IPR000847; HTH\_LysR.  
 CC Pfam: PF00126; HTH\_L1; 1.  
 CC PRINTS: PR00039; HTHLYSR.  
 CC PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
 CC Transcription regulation; Activator; DNA-binding; Plasmid;  
 CC DNA\_BIND 21 40 H-T-II MOTIF (POTENTIAL).  
 CC SEQUENCE 313 AA; 34363 MW; 1B907B3B46B829C7 CRC64;  
 FT  
 SQ  
 Query Match 42.7%; Score 41; DB 1; Length 313;  
 Best Local Similarity 69.2%; Pred. No. 21;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KINLADRMGLGSG 13  
 ||||| |||:|  
 Db 20 KINLAARALGLTG 32  
 RESULT 24  
 CBR2\_RHIME STANDARD; PRT; 313 AA.  
 ID CBR2\_RHIME STANDARD; PRT; 313 AA.  
 AC P56885;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE RUBISCO operon transcriptional regulator.  
 GN CBR.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 ON NCBI\_TaxID=382;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSM419;  
 RA Fennel B.J., Tiwari R.P., Dillworth M.J.;  
 RT "Genetic regulation of C1 metabolism in Sinorhizobium meliloti";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE CBZ OPERON FOR  
 CC RUBISCO AND OTHER CALVIN CYCLE GENES.  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF211846; AAF25374.1; -  
 CC InterPro: IPR000847; HTH\_LysR.  
 CC Pfam: PF00126; HTH\_L1; 1.  
 CC PRINTS: PR00039; HTHLYSR.  
 CC PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
 CC Transcription regulation; Activator; DNA-binding.  
 CC DNA\_BIND 21 40 H-T-II MOTIF (POTENTIAL).  
 CC SEQUENCE 313 AA; 34192 MW; 018CF5EFC0FCED24 CRC64;  
 FT  
 SQ

Query Match 42.7%; Score 41; DB 1; Length 313;  
 Best Local Similarity 69.2%; Pred. No. 21;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KINLADRMGLGSG 13  
 ||||| |||:|  
 Db 20 KINLAARALGLTG 32  
 RESULT 25  
 CD14\_MOUSE STANDARD; PRT; 366 AA.  
 ID CD14\_MOUSE STANDARD; PRT; 366 AA.  
 AC P10810;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Monocyte differentiation antigen CD14 precursor (LPS receptor) (LPS-R)  
 DE (Myeloid cell-specific leucine-rich glycoprotein).  
 GN CD14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 DT MEDLINE=89183627; PubMed=2467257;  
 RA Miyazaki Y., Setoguchi M., Yoshida S., Higuchi Y., Akizuki S.,  
 RA Yamamoto S.;  
 RA "Mouse and human CD14 (myeloid cell-specific leucine-rich  
 RT glycoprotein) primary structure deduced from cDNA clones";  
 RL Biochim. Biophys. Acta 1008:213-222(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR; TISSUE=Macrophage;  
 DT MEDLINE=89287330; PubMed=2472171;  
 RA Setoguchi M., Nasu N., Yoshida S., Higuchi Y., Akizuki S.,  
 RA Yamamoto S.;  
 RA "Mouse and human CD14 (myeloid cell-specific leucine-rich  
 RT glycoprotein) primary structure deduced from cDNA clones";  
 RL Biochim. Biophys. Acta 1008:213-222(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90293480; PubMed=1694207;  
 RA Ferreira E., Hsieh C.T., Franke U., Goyert S.M.;  
 RA "CD14 is a member of the family of leucine-rich proteins and is  
 RT encoded by a gene syntenic with multiple receptor genes";  
 RL J. Immunol. 145:331-336(1990).  
 CC -1- FUNCTION: SERVES AS AN LPS RECEPTOR CONTROLLING CELL ACTIVATION  
 CC UNDER PHYSIOLOGICAL CONDITIONS. WHEN LPS BINDS TO CD14 THE CELLS  
 CC BECOME ACTIVATED AND RELEASE CYTOKINES AND UP-REGULATE CELL  
 CC SURFACE MOLECULES, INCLUDING ADHESION MOLECULES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X13987; CAA32166.1; -  
 CC EMBL: X13333; CAA31710.1; -  
 CC EMBL: M34510; AAA37387.1; -  
 CC PIR: S03605; TDM5M4.  
 CC PIR: A43539; A43539.  
 CC MGD: MGI:88318; CD14.  
 CC InterPro: IPR001611; LRR.  
 CC Pfam: PF00560; LRR; 3.  
 CC PRINTS: PR00019; LEURICHRPT.  
 CC Antigen; Glycoprotein; GPI-anchor; Signal.  
 FT  
 SIGNAL 1 15

FT CHAIN 16 366 MONOCYTE DIFFERENTIATION ANTIGEN CD14.  
 FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 366 AA; 39203 MW; 57C4492EC7EA3AA1 CRC64;

Query Match 42.7%; Score 41; DB 1; Length 366;  
 Best Local Similarity 63.6%; Pred. No. 25;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMLGSGVQEI 17  
 I:|:|:|:|:|:  
 Db 107 RVLGISGLQEL 117

Search completed: June 16, 2002, 00:10:13  
 Job time: 661 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2002, 00:08:55 ; Search time 204.58 Seconds  
(without alignments)  
16.912 Million cell updates/sec

Title: US-09-171-432a-46

Perfect score: 96

Sequence: 1 KINLADRMIGLSCVQEIREQ 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

SPPREMBL\_19: \*  
1: SP\_archaea: \*  
2: SP\_bacteria: \*  
3: SP\_fungi: \*  
4: SP\_human: \*  
5: SP\_invertebrate: \*  
6: SP\_mammal: \*  
7: SP\_mhc: \*  
8: SP\_organelle: \*  
9: SP\_phage: \*  
10: SP\_plant: \*  
11: SP\_rodent: \*  
12: SP\_virus: \*  
13: SP Vertebrate: \*  
14: SP\_unclassified: \*  
15: SP\_rvirus: \*  
16: SP\_bacteriap: \*  
17: SP\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	251	12 09ENR1	09enr1 hepatitis a
2	96	100.0	251	12 09EN09	09eng9 hepatitis a
3	96	100.0	251	12 09EN08	09eng8 hepatitis a
4	96	100.0	251	12 09EN06	09eng6 hepatitis a
5	96	100.0	251	12 09EN05	09eng5 hepatitis a
6	96	100.0	251	12 09EN04	09eng4 hepatitis a
7	96	100.0	251	12 09EN03	09eng3 hepatitis a
8	96	100.0	251	12 09EN02	09eng2 hepatitis a
9	96	100.0	251	12 09EN00	09eng0 hepatitis a
10	96	100.0	251	12 09ENP9	09enp9 hepatitis a
11	96	100.0	251	12 09ENP8	09enp8 hepatitis a
12	96	100.0	251	12 09ENP7	09enp7 hepatitis a
13	96	100.0	251	12 09ENP6	09enp6 hepatitis a
14	96	100.0	251	12 09ENP5	09enp5 hepatitis a
15	96	100.0	251	12 09ENP4	09enp4 hepatitis a
16	96	100.0	251	12 09ENP3	09enp3 hepatitis a

## ALIGNMENTS

17	96	100.0	251	12 09ENP2	09enp2 hepatitis a
18	96	100.0	251	12 09ENP1	09enp1 hepatitis a
19	96	100.0	251	12 09ENN9	09enn9 hepatitis a
20	96	100.0	251	12 09ENN6	09enn6 hepatitis a
21	96	100.0	251	12 09ENN4	09enn4 hepatitis a
22	96	100.0	251	12 09ENN2	09enn2 hepatitis a
23	96	100.0	1124	12 084780	084780 hepatitis a
24	96	100.0	1161	12 005794	005794 hepatitis a
25	96	100.0	2216	12 09WMA2	09wma2 hepatitis a
26	96	100.0	267824	12 067824	067824 hepatitis a
27	96	100.0	2218	12 067817	067817 hepatitis a
28	96	100.0	2225	12 09DL32	09dl32 hepatitis a
29	96	100.0	2227	12 067825	067825 hepatitis a
30	96	100.0	2227	12 067826	067826 hepatitis a
31	96	100.0	2227	12 09WMA4	09wma4 hepatitis a
32	96	100.0	2227	12 09WMA3	09wma3 hepatitis a
33	96	100.0	2227	12 09WMA1	09wma1 hepatitis a
34	96	100.0	2227	12 09WMA0	09wma0 hepatitis a
35	96	100.0	2227	12 09WMA9	09wma9 hepatitis a
36	96	100.0	2227	12 09IFH5	09ifh5 hepatitis a
37	92	95.8	251	12 09ENP0	09enp0 hepatitis a
38	92	95.8	251	12 09ENN7	09enn7 hepatitis a
39	92	95.8	251	12 09ENN5	09enn5 hepatitis a
40	88	91.7	251	12 09ENR0	09enr0 hepatitis a
41	88	91.7	251	12 09ENN8	09enn8 hepatitis a
42	83	86.5	251	12 09EN07	09en07 hepatitis a
43	76	79.2	251	12 09EN01	09en01 hepatitis a
44	72	75.0	251	12 09ENN3	09enn3 hepatitis a
45	71	74.0	184	12 087092	087092 simlan hepa
46	44	45.8	256	16 092SG1	092sg1 rhizobium m
47	44	45.8	414	10 09S248	09s248 arabidopsis
48	44	45.8	788	3 007381	007381 saccharomyc
49	43	44.8	146	5 021062	021062 caenorhabdi
50	43	44.8	211	10 09SEF4	09sef4 arabidopsi
51	43	44.8	232	12 09PYV8	09pyv8 xestia c-nl
52	43	44.8	370	16 067418	067418 aquifex aeo
53	43	44.8	922	12 091LB5	091lb5 white spot
54	42	43.8	104	5 093889	093889 caenorhabdi
55	42	43.8	106	13 098TP6	098tp6 platicthys
56	42	43.8	115	16 09CD22	09cd22 lactococcus
57	42	43.8	374	2 093024	093q24 salmonella
58	42	43.8	414	5 09N691	09n691 euplotes oc
59	42	43.8	528	10 09AU24	09au24 oryza sativ
60	42	43.8	676	16 0985C0	0985c0 rhizobium l

RESULT 1  
ID 09ENR1 PRELIMINARY: PRT: 251 AA.

AC 09ENR1;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE POLYPEPTIDE (FRAGMENT).  
OC Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
OC [1]  
RP SEQUENCE FROM N.A.  
RP STRAIN-A1;  
RA Fujisawa K.;  
RT "Hepatitis A virus.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047652; BAB12160.1; -;  
FT NON TER 1  
FT NON TER 251  
SQ SEQUENCE 251 AA: 28749 MW: 58A520D873893445 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
DB 125 KINLADRMGLSGVOEIKQ 144

## RESULT 2

Q9EN09 PRELIMINARY; PRT; 251 AA.  
AC Q9EN09;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepacovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A159;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047654; BAB12162.1; -.  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
DB 125 KINLADRMGLSGVOEIKQ 144

## RESULT 3

Q9EN08 PRELIMINARY; PRT; 251 AA.  
AC Q9EN08;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepacovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A160;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047655; BAB12163.1; -.  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28693 MW; C32AD5651506751A CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
DB 125 KINLADRMGLSGVOEIKQ 144

RESULT 4  
Q9EN06 PRELIMINARY; PRT; 251 AA.  
AC Q9EN06;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepacovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A162;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047657; BAB12165.1; -.  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28699 MW; 56AA86DF60F6FD3B CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
DB 125 KINLADRMGLSGVOEIKQ 144

## RESULT 5

Q9EN05 PRELIMINARY; PRT; 251 AA.  
AC Q9EN05;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepacovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A20;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047658; BAB12166.1; -.  
FT NON\_TER 1 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20  
DB 125 KINLADRMGLSGVOEIKQ 144

## RESULT 6

Q9EN04 PRELIMINARY; PRT; 251 AA.  
AC Q9EN04;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A201;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047659; BAB12167.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 251 AA; 28720 MW; C334248282F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 125 KINLADRMGLSGVQEIKEQ 144

RESULT 7  
O9EN03 PRELIMINARY; PRT; 251 AA.  
AC O9EN03;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A204;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047660; BAB12168.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 251 AA; 28693 MW; C32AD5651506751A CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 125 KINLADRMGLSGVQEIKEQ 144

RESULT 8  
O9EN02 PRELIMINARY; PRT; 251 AA.  
AC O9EN02;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A205;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047661; BAB12169.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 251 AA; 28723 MW; 558A2D3664C7343C CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 125 KINLADRMGLSGVQEIKEQ 144

RESULT 9  
O9EN00 PRELIMINARY; PRT; 251 AA.  
AC O9EN00;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A302;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047663; BAB12171.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 251 AA; 28644 MW; 1F77AA3ECF3D66B8 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVQEIKEQ 20  
Db 125 KINLADRMGLSGVQEIKEQ 144

RESULT 10  
O9ENP9 PRELIMINARY; PRT; 251 AA.  
AC O9ENP9;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A303;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047664; BAB12172.1; -.

FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28752 MW; 7215A28AD2CA5C1A CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOETKEQ 20  
Db 125 KINLADRMGLSGVOETKEQ 144

## RESULT 11

Q9ENP8 PRELIMINARY; PRT; 251 AA.  
ID 09ENP8;  
AC 09ENP8;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A304;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047665; BAB12173.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28693 MW; C32AD5651506751A CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOETKEQ 20  
Db 125 KINLADRMGLSGVOETKEQ 144

## RESULT 12

Q9ENP7 PRELIMINARY; PRT; 251 AA.  
ID 09ENP7;  
AC 09ENP7;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A306;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047666; BAB12174.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KINLADRMGLSGVOETKEQ 20  
Db 125 KINLADRMGLSGVOETKEQ 144

## RESULT 13

Q9ENP6 PRELIMINARY; PRT; 251 AA.  
ID 09ENP6;  
AC 09ENP6;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A307;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047667; BAB12175.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28644 MW; 1F77AA3ECF3D66B8 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOETKEQ 20  
Db 125 KINLADRMGLSGVOETKEQ 144

## RESULT 14

Q9ENP5 PRELIMINARY; PRT; 251 AA.  
ID 09ENP5;  
AC 09ENP5;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A407;  
RA Fujiwara K.;  
RT "hepatitis A virus";  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047668; BAB12176.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 251 251  
SQ SEQUENCE 251 AA; 28614 MW; 8334EF179C757A6D CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOETKEQ 20  
Db 125 KINLADRMGLSGVOETKEQ 144



```
RESULT 15
ID Q9ENP4 PRELIMINARY; PRT; 251 AA.
AC Q9ENP4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DE 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RC SEQUENCE FROM N.A.
RA Fujiiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047669; BAB12177.1; -.
FT NON_TER 1 1
FT NON_TER 251 251
SQ SEQUENCE 251 AA; 28666 MW; 5A315F46C830D214 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 4,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20
DB 125 KINLADRMGLSGVOEIKQ 144

RESULT 16
ID Q9ENP3 PRELIMINARY; PRT; 251 AA.
AC Q9ENP3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RC SEQUENCE FROM N.A.
RA Fujiiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047670; BAB12178.1; -.
FT NON_TER 1 1
FT NON_TER 251 251
SQ SEQUENCE 251 AA; 28693 MW; C32AD5651506751A CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 4,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20
DB 125 KINLADRMGLSGVOEIKQ 144

RESULT 17
ID Q9ENP2 PRELIMINARY; PRT; 251 AA.
AC Q9ENP2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
```

```
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RC SEQUENCE FROM N.A.
RA Fujiiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047671; BAB12179.1; -.
FT NON_TER 1 1
FT NON_TER 251 251
SQ SEQUENCE 251 AA; 28622 MW; B5C3CD146D39D02 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 4,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20
DB 125 KINLADRMGLSGVOEIKQ 144

RESULT 18
ID Q9ENP1 PRELIMINARY; PRT; 251 AA.
AC Q9ENP1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RC SEQUENCE FROM N.A.
RA Fujiiwara K.;
RT "hepatitis A virus.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047672; BAB12180.1; -.
FT NON_TER 1 1
FT NON_TER 251 251
SQ SEQUENCE 251 AA; 28663 MW; C7EA66FBD19A1619 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;
Best Local Similarity 100.0%; Pred. No. 4,4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVOEIKQ 20
DB 125 KINLADRMGLSGVOEIKQ 144

RESULT 19
ID Q9ENN9 PRELIMINARY; PRT; 251 AA.
AC Q9ENN9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12092;
RN [1]
RC SEQUENCE FROM N.A.
```

RC STRAIN-A68;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB047674; BAB12182.1; -  
FT NON\_TER 1  
SQ SEQUENCE 251 AA; 28666 MW; 5A315F46C830D214 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVGEIKQ 20  
DB 125 KINLADRMGLSGVGEIKQ 144

## RESULT 20

O9ENN6 PRELIMINARY; PRT; 251 AA.  
AC O9ENN6;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A713;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB047677; BAB12185.1; -  
FT NON\_TER 1  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVGEIKQ 20  
DB 125 KINLADRMGLSGVGEIKQ 144

## RESULT 21

O9ENN4 PRELIMINARY; PRT; 251 AA.  
AC O9ENN4;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A77;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB047679; BAB12187.1; -  
FT NON\_TER 1  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

SQ SEQUENCE 251 AA; 28734 MW; 3B96B782882F19D9 CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVGEIKQ 20  
DB 125 KINLADRMGLSGVGEIKQ 144

## RESULT 22

O9ENN2 PRELIMINARY; PRT; 251 AA.  
AC O9ENN2;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A9;  
RA Fujiwara K.;  
RT "hepatitis A virus."  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB047681; BAB12189.1; -  
FT NON\_TER 1  
SQ SEQUENCE 251 AA; 28720 MW; C3342482882F19CA CRC64;

Query Match 100.0%; Score 96; DB 12; Length 251;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVGEIKQ 20  
DB 125 KINLADRMGLSGVGEIKQ 144

## RESULT 23

O84780 PRELIMINARY; PRT; 1124 AA.  
AC O84780;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS) (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,  
RA Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prikhod'ko G.G.,  
RA Bilnov V.M., Vasilchenko S.K., Sandakchiev L.S., Kusov Y.Y.,  
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;  
DR EMBL; X04200; CAA27297.1; -  
DR EMBL; A11312; CAA00953.1; -  
KW Nonstructural protein.  
FT NON\_TER 1  
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

```
Query Match          100.0%; Score 96; DB 12; Length 1124;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVOEIKQ 20
    |||
Db 918 KINLADRMGLSGVOEIKQ 937

RESULT 24
005794 PRELIMINARY; PRT; 1161 AA.
AC 005794; 067800; 067801; 067802; 067803; 067804; 067805; 067806;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P2D (EC 2.7.7.48)) (FRAGMENT).
OS Hepatitis A virus.
OC Hepatitis A virus.
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,
RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,
RA Balayan M.S.;
RL Mcl. Gen. Microbiol. Virol. 6:129-133(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC EMBL; X15464; CAA33492.1; -.
DR Interpro: IPR000408; RCL1.
DR PROSITE: PS00626; RCL1_2; UNKNOWN_1.
KM Polypeptide: Coat protein; Core protein; RNA-directed RNA polymerase;
KM Hydrolase; Thiol protease.
KW CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).
FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 >1161 CORE PROTEIN P2C.
FT NON_TER 1161 1161
SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match          100.0%; Score 96; DB 12; Length 1161;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVOEIKQ 20
    |||
Db 955 KINLADRMGLSGVOEIKQ 974

RESULT 25
09WMA2 PRELIMINARY; PRT; 2216 AA.
AC 09WMA2;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH3;
RX MEDLINE=21386014; Pubmed=11495028;
```

```
RA Fujiwara K., Yokosuka O., Fukai K., Imazeki F., Saisho H., Omata M.;
RT "Analysis of full-length hepatitis A virus genome in sera from
RT patients with fulminant and self-limited acute type A hepatitis.";
RL J. Hepatol. 35:112-119(2001).
DR EMBL; AB020566; BAA35104.1; -.
DR MEROPS; C03.005; -.
DR Interpro: IPR004004; Callci_pol_hel.
DR Interpro: IPR000408; RCL1.
DR Interpro: IPR000605; RNA_helicase.
DR Interpro: IPR001205; RNA_pol_P3D.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
DR PROSITE; PS00626; RCL1_2; UNKNOWN_1.
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match          100.0%; Score 96; DB 12; Length 2216;
Best Local Similarity 100.0%; Pred. No. 5.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KINLADRMGLSGVOEIKQ 20
    |||
Db 961 KINLADRMGLSGVOEIKQ 980
```

Search completed: June 16, 2002, 00:08:55  
Job time: 793 sec

---

**THIS PAGE BLANK (USPTO)**